5-2019

A two-stage approach to ridesharing assignment and auction in a crowdsourcing collaborative transportation platform.

Peiyu Luo
University of Louisville

A TWO-STAGE APPROACH TO RIDESHARING ASSIGNMENT AND AUCTION IN A CROWDSOURCING COLLABORATIVE TRANSPORTATION PLATFORM

By

Peiyu Luo
M.S., University of Louisville, 2013

A Dissertation
Submitted to the Faculty of the
J.B. Speed School of Engineering of the University of Louisville
in Partial Fulfillment of the Requirements
for the Degree of

Doctor of Philosophy
in Industrial Engineering

Department of Industrial Egnineering
University of Louisville
Louisville, Kentucky

May 2019

A TWO-STAGE APPROACH TO RIDESHARING ASSIGNMENT AND AUCTION IN A CROWDSOURCING COLLABORATIVE TRANSPORTATION PLATFORM

By

Peiyu Luo
M.S., University of Louisville, 2013

A Dissertation Approved On

February 7th, 2019

by the following Dissertation Committee

Dr. Lihui Bai, Dissertation Director

Dr. Suraj Alexander

Dr. Kihwan Bae

Dr. Jian Guan

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my dissertation director, Dr. Lihui Bai, for her never ending support and motivation. Her guidance along the way was really valuable and her patience and encouragements never failed to make me feel confident again in the research that I am doing. My sincere thanks go to Dr. Suraj Alexander, Dr. Kihwan Bae and Dr. Jian Guan for reviewing and providing insightful and valuable comments to improve this dissertation.

My greatest appreciation goes to my beloved parents, Mingxia Wang and Zhili Luo for their love and support. Without their trust, I would have never reached this far.

Last but not least, I would like to thank my soul mate and my dear wife Yuan Zhang. Her support means everything to me, and I cant thank her enough for encouraging me throughout this experience at University of Louisville.

ABSTRACT

A TWO-STAGE APPROACH TO RIDESHARING ASSIGNMENT AND AUCTION IN A CROWDSOURCING COLLABORATIVE TRANSPORTATION PLATFORM

Peiyu Luo

February 7th, 2019

Collaborative transportation platforms have emerged as an innovative way for firms and individuals to meet their transportation needs through using services from external profit-seeking drivers. A number of collaborative transportation platforms (such as Uber, Lyft, and MyDHL) arise to facilitate such delivery requests in recent years. A particular collaborative transportation platform usually provides a two sided marketplace with one set of members (service seekers or passengers) posting tasks, and the another set of members (service providers or drivers) accepting on these tasks and providing services. As the collaborative transportation platform attracts more service seekers and providers, the number of open requests at any given time can be large. On the other hand, service providers or drivers often evaluate the first couple of pending requests in deciding which request to participate in. This kind of behavior made by the driver may have potential detrimental implications for all parties involved. First, the drivers typically end up participating in those requests that require longer driving distance for higher profit. Second, the passengers tend to overpay under a competition free environment compared to the situation where the drivers are competing with each other. Lastly, when the drivers

and passengers are not satisfied with their outcomes, they may leave the platforms. Therefore the platform could lose revenues in the short term and market share in the long term. In order to address these concerns, a decision-making support procedure is needed to: (i) provide recommendations for drivers to identify the most preferable requests, (ii) offer reasonable rates to passengers without hurting drivers profit. This dissertation proposes a mathematical modeling approach to address two aspects of the crowdsourcing ridesharing platform. One is of interest to the centralized platform management on the assignment of requests to drivers; and this is done through a multi-criterion many to many assignment optimization. The other is of interest to the decentralized individual drivers on making optimal bid for multiple assigned requests; and this is done through the use of prospect theory. To further validate our proposed collaborative transportation framework, we analyze the taxi yellow cab data collected from New York city in 2017 in both demand and supply perspective. We attempt to examine and understand the collected data to predict Uber-like ridesharing trip demands and driver supplies in order to use these information to the subsequent multi-criterion driver-to-passenger assignment model and driver's prospect maximization model. Particularly regression and time series techniques are used to develop the forecasting models so that centralized module in the platform can predict the ridesharing demands and supply within certain census tracts at a given hour. There are several future research directions along the research stream in this dissertation. First, one could investigate to extend the models to the emerging concept of "Physical Internet" on commodity and goods transportation under the interconnected crowdsourcing platform. In other words, integrate crowdsourcing in prevalent supply chain logistics and transportation. Second, it's interesting to study the effect of Uber-like crowdsourcing transportation platforms on existing traffic flows at the various levels (e.g., urban and regional).

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Travelers select a suitable transportation method (e.g., public transit, private cars, taxi services) considering multiple criteria such as cost, travel time, flexibility, reliability and security. Generally speaking, a public transportation system provides a fixed traveling route and schedule with lower cost and less flexibility. In contrast, private cars or cab services provide faster, more convenient alternatives at much higher cost. Additionally, limited gasoline resource, traffic congestion and green house emission are additional factors that may influence travelers choice of a transportation method towards efficient and sustainable mobility.

Ridesharing refers to the notion where individual drivers provide their personal vehicle and share the travel cost with other participants who have the similar travel itinerary and time schedules. Essentially, ridesharing combines the flexibility and speed of private cars with the reduced cost of fixed-line systems (e.g., public transit) at the expense of convenience. The first regulation of ridesharing in North America can be found through car clubs or car-sharing clubs during World War II [1]. Since then, the usage of ridesharing dramatically increased and reached its peak in 1970 when roughly 21% of American workers commuted to work by carpool. Due to the drop in gasoline price, this rate declined from 19.7% to 13.4% between 1980 and 1990. In 2008, the U.S. Census Bureau report [2] indicates that only 10.7% of the American workers are still using carpool. On the other hand, although considered as an ideal alternative between public and private transportation methods, ridesharing still faces several challenges today. First, imperfect information between drivers and passengers can result in the potential

failure of accommodating unexpected changes of schedule. Second, the general lack of thorough identification of drivers background (e.g., criminal record and driving history) remains to be detrimental to the wide adoption of ridesharing. Third, high transaction cost exists between drivers and passengers, including the time to establish the ridesharing arrangement and the time to pick up and drop off passengers. Other challenges facing ridesharing includes, for example, the reliability of service, the schedule flexibility, consistency of expectations, among others. [3].

On the other hand, the dynamic and real-time ridesharing systems have emerged with the premise of addressing some of the above concerns from the traditional" ridesharing systems. Agatz et al. [4] conclude among several features of a dynamic ridesharing system, one main feature distinguish from traditional ridesharing is that a single, non-recurring ridesharing match could be established on short notice. Moreover, most of the drivers serving in a dynamic ridesharing program are independent and using their personal vehicles rather than centralized employment. To some extent, the success of dynamic crowdsourcing-based ridesharing relies on the development of algorithms for optimally matching drivers and riders, and scheduling and routing participants and vehicles in real-time. The operations research community has only recently started to address the related optimization challenges [5].

In the past decade, rapid developments of mobile computing and social media have elevated dynamic ridesharing to a new plateau, i.e., ridesharing essentially can be seen as crowdsourcing in a collaborative or sharing economy for the transportation system [6]. Companies like Uber and Lyft have grown rapidly by providing so called for-profit-ridesharing platforms via mobile apps for individual users (either service provider or service seeker). Besides the aforementioned differences between dynamic and traditional ridesharing, profit-seeking and flexible schedules are two main features distinguishing Uber or Lyft drivers from those in the traditional ridesharing programs. These platforms mainly use mobile

applications to connect passengers with drivers, who provide ride services using their private vehicles. Passengers are charged with a suggested price based upon a *prior* agreement, where Uber will take a certain percentage as commission fee. The price is usually rather competitive comparing to the rate charged by taxi services. However, as such collaborative platforms receive rapid penetration in the passenger transportation system, the number of passengers requests, which are sent out at any time, can grow rapidly. Therefore, it is important to develop an automatic system for evaluating passenger requests within proper spatial and temporal range, disseminating them to proper drivers and guiding the driver-passenger assignment while respecting drivers autonomous decisions (e.g., whether to serve a passenger, fee charged).

Several attempts have been made (e.g., Ghoseiri et al. [7], Amey [8] and Heinrich [9]) in the literature to address the above issues, although mostly concerning the dynamic or real-time ridesharing system. For example, while many studies focus on automatic ride matching between drivers and passengers, Agatz et al.[10] examine a dynamic model in the Atlanta area that minimizes the total system vehicle miles. Further, Tao et al. [11] develop a greedy method for a dynamic taxi-pooling problem. These instances and most current literature in the dynamic ridesharing research focus on fixed assignments that minimize the system wide measures (e.g., total costs or travelled distances). However, it is envisioned that the distributed nature of the collaborative ridesharing such as Uber and Lyft requires any assignment or pricing decision to respect individual preferences of the participants, and thus a combined centralized and decentralized modeling approach.

In this dissertation, we first propose a multi-criterion optimization based centralized mathematical model to determine the best possible multiple-driver to multiple-passenger matches in an Uber-like platform, with the goal of maximizing the average gender matching rate, average assigned drivers' review rating and the system-wide profit. Based on the initial assignments from the optimization model,

we then develop a decentralized decision making framework to recommend the proper bidding prices for drivers and ultimately leading to one optimal passenger request that the driver would accept in maximizing the prospect of his/her profit. More specifically, the problem is solved through a two-stage process. In the first stage, a many-to-many assignment model is developed to determine tentative matching for multiple passenger requests. In this initial assignment, three criteria important to the Uber-like transportation platform, including gender matching, review rating and total system-wide profit are considered. In the second stage, the Prospect Theory, firstly proposed by Kahneman and Tversky [12], is used to model driver's decision on bidding prices for the multiple passenger requests preliminarily assigned in the first stage as well as the decision on which passenger's request he/she would ultimately take. It is assumed that drivers, faced with possible winning and losing any bid, would evaluate the associated winning and losing probabilities and wish to maximize the prospect of the expected profit.

The remainder of this dissertation is organized as follows. In Chapter 2, we review previous research related to ridesharing problem, web-based logistics, Prospect Theory, many-to-many assignment problem and taxi demand/supply generation. In Chapter 3, we model the multiple drivers to multiple passengers matching as multi-criteria mixed integer program and then propose a prospect theory based decision making framework for drivers to determine bidding price and to select optimal passenger request in Chapter 4. In Chapter 5, data mining and predictive modeling techniques are applied to New York City taxi data to further validate our proposed model, followed by conclusions and future research in Chapter 6.

CHAPTER 2

LITERATURE REVIEW

This chapter includes four sub-sections. In Section 2.1, we review various types of ridesharing problems and their associated models. We then review crowdsourcing-based logistics applications in Section 2.2. Section 2.3 provides an overview on various models and solution algorithms for many-to-many assignment problem. In Section 2.4, the prospect theory and its application relevant to this dissertation are discussed. Finally, we review literature related to taxi demand and supply generation in Section 2.5.

2.1 Ridesharing Systems

Ridesharing provides an alternative means of efficient and environmentally friendly transportation mode in which individual drivers provide their personal vehicle and share the traveling cost with other participants who have the similar travel patterns and time schedules. Essentially, ridesharing is a system that can combine the flexibility and speed of private cars with the reduced cost of fixed-line systems, at the expense of convenience. The social benefit of ridesharing for participants, referring to drivers and passengers, include but is not limited to saving travel cost, reducing travel time, mitigating traffic congestion, and reducing carbon emission. (See Chan et al. [1] for systemic review the history of ridesharing in North American.) In this section, a thorough literature review of the ridesharing problem is conducted to evaluate the previous research work and how they will help to inspire our research. Our review begins with key survey papers on ride-sharing, then

discusses various classification schemes for the ridesharing problem, and finally, proceeds with reviews of individual papers relevant to our research.

The objective of a particular ridesharing system could vary depending on the system operator's motivation, either a private system is driven by maximizing profit or a public system by maximizing social welfare. As Agatz et al. points out in [4], most studies on the ride-sharing problem consider one or more of following objectives:

1. Minimize the system-wide vehicle miles;
2. Minimize the system-wide travel time;
3. Maximize the number of participants.

The above-mentioned objectives can be conflicting at times, thus tradeoff among the maximizing number of participants, minimizing operation costs and minimizing passenger inconvenience need to be made. Further, the ride matching problems often have to include vehicle routing and passenger assignment decisions.

On the other hand, in the survey paper by Furuhata et al. [13], the authors develop a classification system of ridesharing problem from two key aspects: planning and pricing. Planning refers to the successful matching between drivers and passengers with respect to individual preference and time feasibility, while pricing refers to the amount of money transferred between the involved parties in terms of shared cost on gas, toll, parking fees. Furthermore, they identify three other three major challenges for ridesharing agencies, i.e., the design of attractive mechanisms, proper ride arrangement, and building of trust among unknown travelers in online systems. In addition, authors in [14] also present a categorization of ridesharing problems with respect to drivers and riders origins and destinations. Four categories include: identical ridesharing (i.e., the destination and origin for driver and passenger are the same); inclusive ridesharing (i.e., both the origin and destination of the riders are both included in drivers origin to destination path);

partial ridesharing, i.e., refers to the both the pickup location and drop off location of passenger are on the way of drivers origin to destination path, but passengers destination or origin is out of way; detour ridesharing, passengers origin or destination or both of them are not on the way of drivers origin to destination path, therefore, a detour is necessary to meet the passengers request.

The study of systematic investigations on ridesharing has been triggered in the early 1970s. Researchers mainly focus on the following issues: 1. the determinants of mode choice on ridesharing among other modes; 2. switching behavior towards ridesharing; 3. the ride matching optimization. In this section, we present our literature review mainly on issues 1 and 3 and finally summarize the key observations.

Ben-Akiva and Atherton [15] propose a discrete choice model to investigate the switching probability towards ridesharing in the presence of incentive policies. They categorize long range decisions as employment location, residential location, and housing type; medium range decisions as automobile ownership and mode to work; and short range decisions as non-work travel (frequency, destination, and mode), at each level, choices can be modeled using such proposed discrete choice framework. Similarly, Train [16] developes a disaggregate discrete mode choice model that using multinomial logit (MNL) model to study passengers behavior when ridesharing is available as one of the feasible alternatives in the San Francisco Bay Area.

Meanwhile, it is also important to consider the individual ridesharing participants personal preference. For example, through a factorial design survey, Levin [17] conductes two experiments by varying driving arrangement, a size of carpool, distance traveled, and the amount of time to pick up and deliver passengers, to study the quantify attitudes towards alternative carpooling strategies. Their results suggest that the availability of potential ridesharing partners and their relationships could be extremely important among other

considerations. In fact, their findings comply with Ferguson et al.[18], in which, the authors use a confirmatory factor analysis to analyze perceptions towards traditional ridesharing and large scale ridesharing, such as vanpooling. A study of 15 vanpool programs in Southern California operating over 700 vanpools with more than 8,000 members was used to test for such effects. Their results show that reliability and gender are the two factors that have the largest statistical effect.

Furthermore, in transportation economics, the value of time is also identified as a critical variant that affects traveler's switch tendency between drive alone and ridesharing. Huang et al. [19] develope a logit-based stochastic model to investigate how ridesharing is affected by fuel cost, assembly cost, value of time, attitudinal factors, and traffic congestion. It is found that ridesharing is greatly influenced by traffic congestion if a congestion externality-based tolling scheme can be implemented.

In addition, Washbrook et al. [20] study the effects of congestion price under a discrete choice model. Their results suggest that there is a certain level of tolerance to time and cost to switch from drive along to ridesharing, although this tolerance level would vary from person to person and depends on upon the purpose of trip. On the other hand, ride matching formation is an important process, and often the objectives are to minimize the system travel miles or time. For example, Amey [21] study the ride-share problem at the MIT campus in Cambridge, Massachusetts. A data-driven methodology is proposed to estimate the viability of ridesharing, that is, two commuters were arranged to share a trip with given the locations and time constraints among those participants. The ride match process not only determines the driver and passenger pair but also the role (either driver or passenger). A general network flow problem that minimizes the system-wide travel miles was formulated. The results indicate that a potential reduction of system-wide travel mile can be achieved between 9% and 27%, depending on the maximum acceptable driver detour.

Baldacci et al. [22] propose both an exact and heuristic method to solve the ridesharing problem based on two integer programming formulations. Multiple objectives are considered including minimizing vehicle miles and maximizing the number of participants. The passengers are allowed to customize their maximum excess travel time they are willing to accept. Similarly, Calvo [23] studies the problem using a model that allows different network travel times at different times of the day. They develop a heuristic approach to solve the problem.

The major portion of the literature discussed above concentrate on two folds. First, the primary focus of those discrete choice models is on testing the ability of discrete choice models to successfully predict the future impacts in travel behavior rather than evaluate the traveler's travel behavior for a certain purpose, e.g. making profit. Second, the ride matching optimization models primary focus on developing effective algorithm. However, a comprehensive investigation of driver's profit driven motivation in such Uber-like ridesharing platform is absent.

2.1.1 Dial-a-Ride Problem (DARP)

Traditional ridesharing attempts to match requesters and service providers by proximity rather than the exact locations under each partys schedule feasibility. Thus it cannot accommodate unexpected change on either drivers or passengers schedule. In contrast, by advanced notice, dial-a-ride system (DARS) makes door to door delivery service available to those who cannot use public transportation for disability or handicapped upon special requests. In operations research literature, the DARS is modeled mainly through the vehicle routing problem with pickup and delivery time windows and the scheduling problem specify their preferred pick up and drop off locations between origins and destinations. The problem is shown to be NP-hard [24]. Several variants of the DARS exist, most of which deal with either developing an efficient system to maximize the number of passengers served or minimize passengers waiting time or routing costs. The early work of the DARP can

be found in [25] and [26], where the single vehicle pick-up and delivery problem can be modeled as traveling salesman problem with time window, along with additional constraints of capacity and precedence. Psaraftis [27] proposes an exact dynamic programming algorithm for the single vehicle DARP, to minimize the total customer inconvenience. Its adaptation to the dynamic case including an additional constraint regarding the maximum position shift of the customers is also discussed. Due to the computational time complexity ($n^2 3^n$, where n is the number of requests received) of the backward recursion programming approach, the problem size that can be handled is relatively limited (up to 10 customer requests). A revised version of the dynamic algorithm is presented in [28], where the forward recursion programming approach is used instead of backward method.

2.1.2 Dynamic Ridesharing Problem

By effectively using emerging internet and global positioning system enhanced mobile devices (e.g., smartphone applications), it is possible to allow users (either service requester or service providers) to input their travel information including their origin, destination, time restrictions, expected cost (charge) and so on. Different from traditional ridesharing system or dial-a-ride system, dynamic ridesharing system refers to a system consisting of independent drivers that process the matching between requesters and providers automatically with a very short notice. Unlike traditional carpooling or vanpooling, which usually commit a long-term partnership between participants, dynamic ridesharing focus on one-time, non-recurring trips. One main challenge in dynamic ridesharing problem is that service requesters and providers are entering the process continuously, thus it is necessary to identify the relevant requests and offers before any planning or scheduling is executed.

Dial [29] presents an automatic process for dealing with a DARS with multiple taxi vehicles that allow requesters send their requests with a short notice

via telephone or internet. Using the dynamic programming approach in [28], the new requests were inserted into the tentative optimal routes and then updated by selecting the schedules with minimum insertion cost.

In order to deal with real-time dynamics, one can always solve the static problem iteratively in a rolling horizon framework. For instance, Yang et al. [30] studies a real-time multi-vehicle truckload pickup and delivery problem. In their problem, the trucks dynamically move from site to site, according to customer's dynamic requests. The authors first modeled the static problem as a mixed integer programming, and then the model was solved repeatedly under five rolling horizon. Simulation results were reported in order to evaluate different scheme.

Similarly, Agatz et al. [10] presents an optimization based approach that minimizes the total system-wide travel miles in a dynamic ridesharing system. The weight of the edges that link each driver and passenger are computed based on the total vehicle kilometers of travel (VKT) savings. Two commitment strategies were considered, immediately notice for the drivers whenever the requests were received and delayed notice until next execution time. The simulation results indicate that delayed notice commitment strategy would produce optimality more frequently than immediately notice commitment strategy does. However, the benefit of immediately notice commitment strategy is allowing the accumulation of more trips announcements between each optimization rolling horizon, thus the utilization of each vehicle increases accordingly.

While most approaches are implemented as classical centralized optimization problems, many researchers also seek for decentralized approaches in order to solve large-scale realistic problem with reduced computation time, such as agent-based modeling. The key concept of decentralized models is that autonomous riders and driver agents are able to establish ridesharing matches locally instead of a centralized collective platform, through wireless sensor networks. One challenge of these agent-based models is the way to disseminate relevant geospatial information

to spatially dispersed mobile users, especially for those new entities entering the system.

In [31], Nittel et al. formally classify the information dissemination strategies in mobile geosensor networks as:

1. Flooding: Whenever a new request received by an agent, he/she will spread this request to other agents within the radio range, and those agents will also pass on the information until every node in this network is notified.

2. Epidemic: the agents will only pass on the new requests to a pre-determined k agents in the network.

3. Location-constrained: requests will be shared with a certain spatially range, any other agent beyond this range will be ignored.

Winter and Nittel [32] study the impact of short range communication device on disseminating users' information under different information dissemination. Their simulation results indicate without centralized planning tool, the average client travel time decreases as the number of host increases. This implied that the quality of the solution is guaranteed. Similarly, Xing et al. [33] consider a highly dynamic ridesharing system where passenger agents seek potential drivers in the network every two minutes. Again, without a centralized planning, individual users announce their trip information at the departure time. A maximum acceptable service response time for the riders are provided, as well as personal preferences (e.g., gender, smoking habit). Simulation results for the Bremen metropolitan area suggested that the sufficient number of drivers promise a higher successful ridesharing matching rate. However, the drawback for both models is that the number of ride matching in the system is not maximized.

In addition, Kleiner et al. [34] propose an auction-based scheme for the dynamic ridesharing with one driver to one passenger setting. They apply a rolling horizon decision-making approach that all assignments were committed by a

deadline. In their simulation, each passenger is willing to pay per mile lies between the cost of private car and the cost of taxi service. Second-price auction scheme is employed to encourage passengers to bid for higher ranking. The simulation experiments show that the auction-based approach provides close-to-optimal solutions to the ride-sharing problem.

2.2 Web-Based Logistics and Crowdsourcing Services

The introduction of crowdsourcing offers numerous business opportunities. In recent years, manifold forms of crowdsourcing have emerged on the market, as well as in logistics. The concepts of crowdsourcing have been known for years. Recent interest in the crowdsourcing platform is carried out by Howe [35]. Crowdsourcing was defined as: the act of a company or institution taking a function once performed by employees and outsourcing it to an undefined (and generally large) network of people in the form of an open call. Crowdsourcing is the platform which mainly focuses on building a network of collaborators and facilitating online communication between various groups, people, organizations, with the similar interest and motivation to help to solve an assigned problem collaboratively. Saxton [36] characterizes crowdsourcing should have following features: the process of outsourcing the problem, the crowd, and a web-based platform for collaboration. In general, outsourcing a problem usually occurs when there is no in-house solution available or it is uneconomic means to produce solution. Relying on vast motivated individuals, the capability of providing solutions superior in quality and quantity is guaranteed to those traditional business modes. Currently, a large number of crowdsourcing based business applications have been successfully stimulated research work within the field of management science. Brabham [37] investigates several notable models including Threadless, iStockphoto, InnoCentive, etc. from their theoretical grounding and realistic cases.

In [6], Alt et al. investigate the possibility of location and context based

crowdsourcing to distribute delivery tasks to motivated service providers. They focus on how potential solution provider reacts to a task via effectively use the new mobile technology and global positioning system. A crowdsourcing platform was implemented that integrates location as an input parameter for distributing tasks to solvers. In general, a vast number of drivers spend hours traveling on the same routes at about the same time on their commute to and from work. Their travel patterns are regulating a highly predictable transportation network. On the other hand, various types of shipping are placed when customers place requests to a shipment carrier to transport an item from its origin location to a destination for a particular fee. The fee charged to a customer often depends on the size and weight of the item being shipped, the distance between the origin and destination, and the shipping speed and time required by the user. Accordingly, there is an opportunity to form a system that is able to provide accessible and affordable services, even allow pick-up and drop-off in rural areas that traditional shipment carriers have no access to. Today people can easily interact and share information with each other with the popular use of smartphones and other portable devices. Thus it is realistic to enable Crowdsourcing to address the diversity of problems including real-time data collection and processing, dynamic re-routing and cooperation among a large group of participants.

Some researchers have studied the logistics aspects related to crowdsourcing. Carbone et al. [38] identify and describe four types of collaborative logistics: peer-to-peer logistics, business logistics, crowd-party logistics and crowd-driven logistics, via their management types and the role logistics played. Peer-to-peer logistics and business logistics are the frameworks where individuals trade, donate, exchange goods or services based on mutual agreement, arrange and operation of the logistics themselves. The difference between peer-to-peer logistics and business logistics is the collaborative platform plays both as an informational intermediary to share information and as a physical intermediary to perform necessary logistics. On

the other hand, crowd-party logistics and crowd-driven logistics are different from previous two cases where logistics served merely to support the cooperation between peer to peer (manufacturer to customer). In contrast, logistics is the starting point to motivate collaborative arrangement. In these two case, crowd-party logistics mainly focus on tapping into the logistic capacity of the crowd by efficient and economical means, such as Uber, DHL's MyWays service, while crowd-driven logistics adopts a centralized fashion in order to establish a direct contact with the producers, AMAP in France is requiring members to participate the weekly distribution.

Lee et al. [39] present an integrated decision-making framework for on-demand crowdsourcing delivery services that considers Just-In-Time delivery, fuel consumption, and carbon emissions. Based on a continuous variable feedback control, the integrated framework allows unified processing of delivery requests and route scheduling. The computational results show that increase revenue by 6.4% by reducing fuel and emission costs by 2.5%.

To date, although most of the research in the field of crowdsourcing still focus on virtual tasks and its related area, such as design, translation [40]. Some researchers studied the possibility of using crowdsourcing platform to conduct physical task such as package delivery. For example, Suh et al. [41] study the impact of customers share their spatial and networked information from social network, e.g. Facebook, to complete last-mile package delivery systems from online purchases. Rougs et al. rouges2014crowdsourcing propose a Physical Internet-based crowdsourcing delivery model, it enables that each single crowdsourced route to becoming a consolidated segment of a long haul task. Furthermore, some retailer has already triggered the crowdsourced delivery in practice. Walmart is running a so-called "Walmart To Go" same-day delivery program, those in-store customers who accept to deliver packages on their way home to online customers will receive a discount on their purchases [42].

Recently, Arslan et al. [43] study a crowdsourced delivery platform which enables dynamic pick up and delivery with ad-hoc drivers. A job matching model is formulated to in the first stage. However, since one particular job may contain more than one task, therefore, the driver's routing problem become the traveling salesman problem with time window in the second stage. Finally, a rolling horizon approach is used to handle dynamics whenever a new task or driver arrives. The experiment results suggest that using ad-hoc drivers has the potential to reduce the last-mile cost and system-wide vehicle miles.

Meanwhile, several potential issues have been discussed from the strategic perspective. To name a few, several surveys conducted by Watkins et al. [44], indicate personalized, reliable, and up-to-date information have the highest priority to participants. Also, Filippi et al [45] point out the participants should be empowered to influence the logistic service, which will give flexibility to the system and thus foster bottom-up development.

However, most of the current literature focus on the functionalities of the applications, such as UbiGreen, GreenGPS and so on, or explore the sustainability for environmental or economical perspective. There are various motivations can drive the users to participate sharing service including frugality, opportunism, eco-responsibility, etc. Hence, it is lacking detailed crowdsourcing based mathematical models from the operational level in the literature.

2.3 Many-to-many Assignment Problem

Meanwhile, it is also important to consider individual participants personal preference when the centralized platform make tentative assignment between service seeker and service provider. For example, through a factorial design survey, Levin [17] conducts two experiments by varying driving arrangement (e.g., size of carpool, distance traveled, and the amount of times to pick up and deliver passengers), to study and quantify attitudes towards alternative carpooling strategies. Their results

suggest that the availability of potential ridesharing partners and their relationships could be extremely important among other considerations. In fact, their findings comply with Ferguson et al. [18], in which, the authors use a confirmatory factor analysis to analyze perceptions towards traditional ridesharing and large scale ridesharing, such as vanpooling. A study of 15 vanpool programs in Southern California operating over 700 vanpools with more than 8,000 members was used to test for such effects. Their results show that reliability and gender are the two factors that have the largest statistical effect. Therefore, it is envisioned that the distributed nature of the collaborative ridesharing such as Uber and Lyft requires any assignment or pricing decisions to respect multiple individual preferences of the participants, and thus a multi-criterion modeling approach may be applied.

When considering multiple individual preferences, it makes sense to provide flexibility for either driver or passengers in earlier stage so that when multiple criteria are applied in sequential or other manners, optimal solutions are still of high practical quality. Therefore, we consider a two-stage process. In the first stage, we assign passenger requests to drivers allowing for one passengers be preliminarily assigned to more than one drivers, the vice versa. This type of many-to-many assignment problem provides flexibility in solutions that will be passed onto the next stage. In the second stage, each driver will solve their own secondary optimization problem to ultimately choose a passenger request that maximizes his/her own interest.

Although the assignment problem is one of the fundamental combinatorial optimization problems in optimization or operations research of mathematics. For example, one classical solution to the assignment problem is given by the Kuhn-Munkres algorithm, originally proposed by H. W. Kuhn [46] and refined by J. Munkres [47]. The Kuhn-Munkers algorithm assume the a priori existence of a matrix of edge weights, w_{ij} , or costs, c_{ij} and the problem is solved with respect to these values. It is able to solve the assignment problem in $O(n^3)$ time, where n is

the size of one partition of the bipartite graph. However, limited works can be found in literature related to the many-to-many assignment problem.

For instance, Psaraftis [27] proposes an exact dynamic programming algorithm for the single vehicle many to many assignment problem, to minimize the total customer inconvenience. It adapts to the dynamic case by adding a constraint on the maximum position shift of a customer. Due to the computational complexity of the backward recursion approach, the problem size that can be handled is relatively limited. A revised version of the dynamic algorithm is presented in [28], where the forward recursion is used instead of backward method.

More recently, Zhu et al. [48] propose a solution method for solving the many-to-many problem by improving the Kuhn-Munkers algorithm with backtracking. Similarly, in [49] and [50], Litvinchev et al. study a Lagrangian based heuristic for many-to-many assignment problems taking into account capacity limits for task and agents. Based on modified Lagrangian bounds, the authors propose a greedy heuristic to get the Lagrangian-based solution for the many-to-many assignment problem. The greedy heuristic algorithm is also used to speed up the subgradient scheme to solve the modified Lagrangian dual problem.

Durfee et al.[51] propose a new formulation to assign multiple experts to multiple teams, while considering experts' availability as schedule constraints.Their work demonstrates the significance and complexity of the problem of assignment and scheduling between experts and teams. They reformulate the hybrid model into an integer linear programming problem and thus it is can be solved by a standard mathematical programming package.

Further, the study on the field of multi-criterion optimization is quite extensive. Several comprehensive references to multi-criterion optimization could be found in Hwang et al. [52], Ringuest [53] and Steuer [54], and with respect to applications of engineering design in Eschenauer et al. [55] and Anderson [56]. We only concentrate on the multi-criterion optimization in task assignment for the

literature review as it relates to our research.

In general, the multi-criterion optimization can be solved in three different ways towards decision maker's preference on the objectives: priori, posteriori and interactive.

In priori methods, the decision maker is supposed to be aware their preferences clearly by recognizing the significance or weights to the objective functions. For example, the weighted sum and lexicographic approaches are examples of priori methods.

Alternatively, the decision maker progressively gives preference toward the most preferred solution in interactive methods.The decision process converges to the most preferred solution by evaluating solutions iteratively until the decision maker is satisfied with the solution.

In posteriori methods, the set of potential solutions are generated, and later the decision maker selects one among them based on preference. The decision process is divided into two independent phases: the first phase generates all the possible alternatives and the second phase selects the most preferred one among them when all possible choices are available. The ϵ-constraint method and genetic algorithm are commonly used for posteriori methods. [57]

2.4 Prospect Theory

The field of decision making under risk or uncertainty has been studied for years, since von Neumann and Moregenstern [58] firstly propose the expected utility model. However, several experiments have shown that the expected utility model can result in inconsistency between observed choice and the predicted choice under expected theory, i.e., the Allais paradox [51].

Two main approaches can be found in the literature to address such violations. Hey and Orme [59] use the random utility model, which can provide a satisfactory prediction on individual choice by adding the error term into expected

utility model. Since these assumed distributions of error terms can significantly influent the final recommendations on descriptive power of probability choice, thus random utility model need reliable and efficient estimation on decision makers preferences.

The second approach is introduced by Kahneman and Tversky [53], they developed the prospect theory and cumulative prospect theory that explain the paradoxes and some other issues surrounding expect utility theory. Prospect theory posits the decision maker's evaluation in terms of deviation from a reference point instead of a net wealth level, the utility of an outcome is weighted by the weight of the probability instead of the probability of its occurrence. In order to assess the utility of a gain and disutility of an equivalent loss, a risk coefficient is associated with the loss term and the decision makers are usually risk averse over gains but risk seeking over losses.

To our best knowledge, prospect theory is the first and foremost model of decision making under risk, numerous applications can be found within the field that associated with probabilistic outcomes, especially for alternatives under risk, such as finance and insurance, where decision makers attitude towards risk play a central role. For example, in [60], Barberis and Huang propose a prospect theory-based model that evaluate the investor's asset prices over the periods when the investor changes the portion of their investment portfolios. Their results suggest that, under prospect theory prediction, since those stocks are strictly positively skewed (profitable), those investors are willing to pay a high price for the stock, even when it means earning a low average return on it.

Insurance is another interesting field of decision making that applying prospect theory to explain decision makers behavior. A case study of 50,000 customers from a home insurance company can be found in Sydnor [61]. Those customers are willing to pay a higher premium policy with a lower deductible, even the average annual claim rate is extremely low. This is because customers tend to

overweight those risky unpleasant outcomes, even with a relatively low possibility, under prospect theory explanation.

Camerer et al. [62] note that prospect theory can help to understand how labor supply reacts to salary. Using the data from cab drivers in New York City, the authors find that the number of hours that a driver works on a given day is strongly inversely related to his average hourly wage on that day. That is, a particular driver derives prospect theory utility from the difference between his daily income and a certain target level. However, this driver with these preferences will stop work for the day after reaching his target income level.

2.5 Taxi Demand/Supply Generation

In economics, demand refers to the quantity of a product or service that is desired by buyers while supply refers to the quantity of a product or service that is provided by the sellers. Market equilibrium is reached once demand is fulfilled by supply. The taxicab has been the dominant transportation mode in most urban areas all over the world. By governmental regulations of taxi licenses, the taxicab market is then able to combat oversupply and then provides quality service, competitive price and ensures the safety. However,the taxi market is currently facing challenge from rapid growth of ride-sharing applications such as Uber and Lyft [63]. Since the inception of Uber in 2009, it has gained the significant market share from the taxi market in the United States as shown in Figure 1. In 2015, Uber's market share exceed the traditional taxi cab market share for the first time [64].

There exists significant intersection between taxi cabs' customers and crowdsourcing ridesharing customers. Further, the number of potential customers will be systematically analyzed by the characteristics of the target traffic zones, as well as the pattern of pickups and drop-offs in different time of the day.

Understanding how transportation trips are generated and distributed by time and location is very insightful for the centralized policy makers to provide

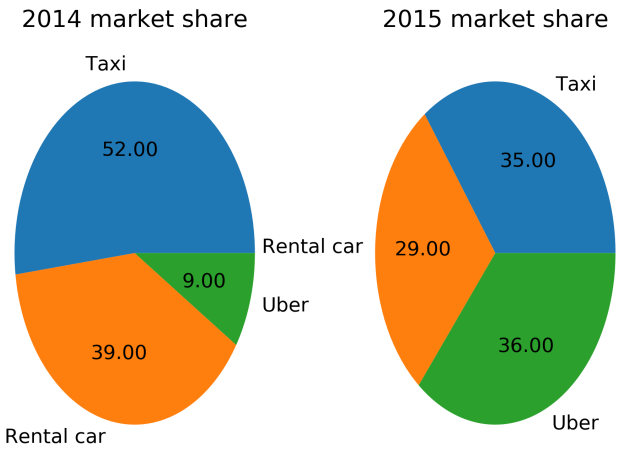

Figure 1. Transportation market share in 2014/2015

transportation service in an efficient way. In the literature, the field related to transportation demand generation is quite extensive. Garber and Hoel [65] defines four steps to model large scale travel demand including: trip generation, trip distribution, modal split and trip assignment. In this dissertation, we focus on the trip generation and trip assignment. We also conduct additional analysis including the empirical parameters estimation (i.e. dollar per mile/trip), and supply forecasting to allocate the potential service providers for generated traffic demand in the nearby areas.

Many of the literature contributes to the policies and regulations on taxi market, especially in the early stage. We only name a few for reference since the taxi regulation and policy making is not our focus in this dissertation. It is interesting to note that public debate surrounding deregulated/regulated taxi market has never reached consensus among economists. To our best knowledge, Turvey [66] may be the first systematical study on the economic features of the taxi market control and regulation in London. There are numerous following case studies in the related theory and discussions. Some researchers such as Coffman and Shreiber [67] discuss the excessively high price scheme in unregulated taxi market if the price information is scarce and search costs high. Dempsey [68] claims that unregulated taxi market

may led to declining fares, long hours for drivers, dangerous cars, and inadequate compensation for accident victims. On the other hand, Gaunt [69] and Marell [70] study the impact of taxi deregulation in New Zealand and Sweden. Their results suggests significant new entry to taxi market and thus lead to quality service, such as fare reductions and less passenger waiting time.

On the other hand, Global Positioning System (GPS) devices enable vehicle-movement tracking ability in recent years, and thus making it easy to monitor traffic and gather valuable geographic data. These data are collected from governmental agencies as well as crowdsourcing services such as Uber and Lyft. With the availability of those data, it is possible to provide empirical analysis of taxi demand in a large urban network. For example, New York City Taxi Limousine Commission has been collecting GPS data for those licensed taxi cabs since 2008. Unfortunately, by the time of this dissertation, Uber only reveals partial trip information (pickup date time, pickup location and dispatch base) in selected cities such as Boston. In fact, it is important for taxi agencies and commissions as well as crowdsourcing companies to share the data with the public in order to help develop some empirical models for taxi demand. And the insights and knowledge of these models can also potentially be used to promoting policies and regulations.The details of theoretical models will be discussed in the following section.

There are three commonly methods used to model trip generation, i.e., rate method, cross classification method and regression method. The rate method is used for traffic impact analysis on non-residential trip generation, which does not consider characteristics such as household size, income, and auto ownership. A cross classification model cross-tabulates average trip making rates with two or more variables, allowing for a clear understanding of important factors without assuming that the relationship between demands and explanatory factors follows a specific functional form or that there is independence between these factors [71].In this dissertation since our approach is using regression models, we will focus on

literature that using regression techniques. Regression model has been a widely used statistical tool to explore the relationship between response variable and explanatory variables. If one can collect enough information, the regression models can be very useful to forecast and analyze the travel demand in urban transportation systems [72]. For instance, if the information related to carpool is available, we may then be able to forecast the demand for carpool. However, in order to obtain quality estimation of target variables, it usually requires a sufficient dataset with detailed information with respect to the corresponding response.

Schaller [73] has performed an analysis on the total number of taxicabs in 118 U.S. cities using multiple linear regression. Schaller uses independent variables such as population, employment,use of complements to taxi cabs (e.g., public transit), cost of taxi, and taxi service quality to predict the number of taxicabs in the fleet as the dependent variable instead of the number of taxi trips generated. Nevertheless Schaller has identified that the influential factors include the number of workers commuting by subway, the number of households with no vehicles available, and the number of airport taxi trips. On the other hand, Mousavi et al. [74] have stated that household structure, age, gender, marital status, income, employment, car ownership, population density, and distance to transit are the most influential variables on trip generation for all modes.

Characteristics of the trip (e.g., travel purpose) and characteristics of the traveler (e.g., age, income, education) have been identified as influential factors affecting the trips generated for different travel modes ([71],[75], [76]). Trips to residential areas and non-residential areas [76] and trips for business and non-business purposes [72] are analyzed separately in some studies. A number of studies have been conducted about trips generated to airports (e.g., [76]) and travel to schools (e.g., [77],[78]).

In this dissertation, we will use selected characteristics of the census tracts to gain insights on the types of people and activities that are most associated with taxi

trip making and we also focuses on the characteristics of the people who live and work in these places in order to develop forecasting models for taxi trip demand.

CHAPTER 3

MANY-TO-MANY ASSIGNMENT PROBLEM USING MULTI-CRITERIA OPTIMIZATION

3.1 Problem Statement

In this chapter, we focus on developing a multi-criterion assignment model to determine the tentative matching where a passenger request is assigned to multiple drivers and a driver is assigned to multiple passenger requests, subject to various constraints such as traveling information, participant's personal information including driver's review rating, gender of drivers and passengers. Three criteria are taken into account. First, in order to address female passenger's safety concern with male non-acquaintances, the female passenger's requests will be assigned to female driver as many as possible. Second, by maximizing the assigned driver's review rating, the centralized platform is able to motivate drivers to provide consistent service quality to earn more business opportunity in the future. Finally, the driver's profit is to be maximized as the decision support systems respects the decentralized decision making.

3.2 Problem Formulation

In this section, we propose a multi-criterion optimization model to address the task assignment problem under many-driver-to-many-passenger circumstance. That is, a particular driver is allowed to receive more than one passenger request, while any one of the passenger's request can be presented to more than one driver. Throughout this dissertation, we refer this type of assignment as many-to-many

assignment. Prior to solving the problem, the information regarding passenger's gender, current location and destination, driver's gender, current location and review rating, the estimated traveling cost under current traffic condition and expected profit from this trip are obtained. Thus, three criteria include: the average gender matching rate, the average driver's review rating and potential profit. Given a set of drivers I and a set of passengers J, let x_{ij} be a binary decision variable, which equals one if the passenger $j \in J$ is assigned to driver $i \in I$ and zero otherwise. The average gender matching can be calculated as follows:

$$F_1 = \frac{\sum_{i,j} g_{ij} x_{ij}}{\sum_{i,j} x_{ij}}, \tag{1}$$

where g_{ij} is a binary indicator which equals 1 if both driver i and passenger j are female and 0 otherwise.

Similarly, associated with the scale from 0 to 5 for driver's review rating r_i, the average assigned driver's rating is formulated as:

$$F_2 = \frac{\sum_{i,j} r_j x_{ij}}{\sum_{i,j} x_{ij}}, \tag{2}$$

Last, the total potential profit is defined in 3 and p_{ij} is the expected profit from driver and passenger pair p_{ij}

$$F_3 = \sum_{i,j} p_{ij} x_{ij} \ . \tag{3}$$

Those above mentioned objective functions are maximized in order to: i) pair female passengers with female drivers as much as possible. ii) give highly reviewed drivers priorit. iii) maximize total profit. Constraints for those objective functions

are listing below:

$$\sum_{i} x_{i,j} \leq \bar{P} \quad \forall j \in J \tag{4}$$

$$\sum_{j} x_{i,j} \leq \bar{D} \quad \forall i \in I \tag{5}$$

$$\sum_{i} x_{i,j} \geq 1 \quad \forall j \in J \tag{6}$$

$$x_{i,j} \in \{0, 1\}, \tag{7}$$

constraint (4) and (5) limit each passenger's request can be assigned to no more than $\bar{P}$ drivers, , as well as the maximum number of requests that one driver can review is also capped and each drivers can only be assigned to no more than $\bar{D}$ passengers. Finally, constraint 6 ensures that each passenger's request can be received by at least one driver. An important feature of this formulation is that it cannot be solved as an integer linear programming model, as objective function F_1 and F_2 are nonlinear.

3.3 Solution Methods

Finally, the multi-criterion many-to-many passenger-driver assignment problem is presented below.

(MCMMAP)

$$
\begin{aligned}
\text{MAX} \quad & F_1 = \frac{\sum_{i,j} g_{ij} x_{ij}}{\sum_{i,j} x_{ij}} \\
\text{MAX} \quad & F_2 = \frac{\sum_{i,j} r_j x_{ij}}{\sum_{i,j} x_{ij}} \\
\text{MAX} \quad & F_3 = \sum_{i,j} p_{ij} x_{ij} \\
\text{st.} \quad & \sum_{i} x_{i,j} \leq \bar{P} \quad \forall j \in J \\
& \sum_{j} x_{i,j} \leq \bar{D} \quad \forall i \in I \\
& \sum_{i} x_{i,j} \geq 1 \quad \forall j \in J \\
& x_{i,j}, \in \{0, 1\},
\end{aligned}
$$

the above formulated MCMMAP essentially is a multi-criterion integer nonlinear program (MCINP), thus is difficult to solve for global optimal solution. In ths section, we propose a novel approach to reformulate the mixed integer non-linear programming under multi-criterion into an equivalent mixed-integer linear programming form. Our approach is based on the reformulation of the denominator of objective function F_1 and F_2, and we explain the details in the following subsections.

3.3.1 Linearized model

We first consider the integer nonlinear program with the objective function F_1 only, denoted as problem (P_1) below:

$$
\begin{aligned}
(P_1)\quad & \\
\text{Max} \quad & F_1 \\
\text{st.} \quad & \sum_i x_{i,j} \le \bar{P} \quad \forall j \in J \\
& \sum_j x_{i,j} \le \bar{D} \quad \forall i \in I \\
& \sum_i x_{i,j} \ge 1 \quad \forall j \in J \\
& x_{i,j} \in \{0,1\},
\end{aligned}
$$

We then introduce a new sequential integer set parameter in ascending order, $k_1 = \{1, 2, ..., K\}$, where K is a sufficient large integer such that $K \ge \sum_{i,j} x_{i,j}$. Therefore, one can always find a k_1^* among the series such that $k_1^* = \sum_{i,j} x_{i,j}$. By introducing an auxiliary variable Y_1, the above model can be linearized as:

$$
\begin{aligned}
\text{Max} \quad & Y_1 & \\
\text{st.} \quad & Y_1 \le \frac{\sum_{i,j} g_{ij} x_{ij}}{k_1} + M(1 - z_{k1}) \quad \forall k_1 = 1, 2, ..., K & (8) \\
& \sum_{i,j} x_{i,j} = \sum_{k_1} z_{k1} k_1 & (9) \\
& \sum_{k_1} z_{k1} = 1 & (10) \\
& \sum_i x_{i,j} \le \bar{P} \quad \forall j \in J & (11) \\
& \sum_j x_{i,j} \le \bar{D} \quad \forall i \in I & (12) \\
& \sum_i x_{i,j} \ge 1 \quad \forall j \in J & (13) \\
& x_{i,j}, z_{k1} \in \{0,1\}, & (14)
\end{aligned}
$$

In constraint (15), M is a sufficient large number and z_{k1} is a binary decision variable to determine the optimal value of k_1. Constraint (16) restrict the total

number of assignment is equal to the special number k_1 and constraint (16) ensure that only the optimal k_1^* can be selected.

Similarly, we linearize objective function F_2 and F_3, thus we have following P2 and P3:

$$\begin{aligned}
\text{Max} \quad & Y_2 \\
\text{st.} \quad & Y_2 \leq \tfrac{\sum_{i,j} r_i x_{ij}}{k_2} + M(1 - z_{k2}) \quad \forall k_2 = 1, 2, ..., K && (15)\\
& \sum_{i,j} x_{i,j} = \sum_{k_2} z_{k2} k_2 && (16)\\
& \sum_{k_2} z_{k2} = 1 && (17)\\
& \sum_{i} x_{i,j} \leq \bar{P} \quad \forall j \in J && (18)\\
& \sum_{j} x_{i,j} \leq \bar{D} \quad \forall i \in I && (19)\\
& \sum_{i} x_{i,j} \geq 1 \quad \forall j \in J && (20)\\
& x_{i,j}, z_{k2} \in \{0, 1\}, && (21)
\end{aligned}$$

$$\begin{aligned}
\text{Max} \quad & Y_3 \\
\text{st.} \quad & Y_2 \leq \sum_{i,j} p_{ij} x_{ij} + M(1 - z_{k3}) \quad \forall k_3 = 1...K && (22)\\
& \sum_{i,j} x_{i,j} = \sum_{k_3} z_{k3} k_3 \quad \forall k_3 = 1...K && (23)\\
& \sum_{k_3} z_{k3} = 1 \quad \forall k_3 = 1...K && (24)\\
& \sum_{i} x_{i,j} \leq N \quad \forall j \in J && (25)\\
& \sum_{j} x_{i,j} \leq N' \quad \forall i \in I && (26)\\
& \sum_{i} x_{i,j} \geq 1 \quad \forall j \in J && (27)\\
& z_{k3} \in \{0, 1\}. && (28)
\end{aligned}$$

3.3.2 Lexicographic solution method

In multi-criterion optimization,if the multiple, say K, objectives follow a dominance order, i.e., decision maker is able to identify if the objective k is of the

highest priority or significance and whether k should be optimized first, before considering the value of rest objective 1, 2, ..., k-1. Practically, the lexicographical optimization is performed as follows. First, the objective of highest priority is solved, obtaining z_1^*. Then the objective of second highest priority is optimized by adding the constraint $z_1 = z_1^*$ in order to remain the optimality of the first objective. With the obtained optimal solution $z_2 = z_2^*$, subsequently, the following objective function is then optimized by adding the constraints $z_1 = z_1^*$ and $z_2 = z_2^*$, the procedure is repeated until all objective functions are solved. In summary, the algorithm is given as:

---

for j:=1 to K :
    Solve max z_j $|x \in X$
    add constraint $z_j = z_j^*$
end for

The solution procedure of our problem by using lexicographical optimization can be summarized in the following step:

Step 1: Re-formulate the model as described in $P1$, $P2$ and $P3$.

Step 2: Solve the first objective function as single objective problem and obtain $Y_1 = Y_1^*$.

Step 3: Solve the second objective function as single objective problem by adding $Y_1 = Y_1^*$ and output $Y_2 = Y_2^*$.

Step 4: Solving the last objective function as single objective problem by adding $Y_1 = Y_1^*$ and $Y_2 = Y_2^*$.

3.4 Computational Results

In this section, we report the results from our numerical tests which are implemented and solved in GAMS [79], a state-of-the-art modeling language for nonlinear programs. CPLEX is employed as the solver for the optimization problem. All instances are run on a 16-core dual Opteron CPU server with 32GB of memory running openSUSE 11 Linux.

3.4.1 Random instance generation

We first discuss how random test instances are generated. There are several parameters that may affect our model's behavior. They include: the ratio between passengers and drivers, the maximum allowed request per driver and the maximum allowed driver per request. The experiment designed here is to test the effects of these parameters on the final solution. The values and ranges of these parameters in the random instances are listed in Table 1.

Without loss of generality, we assume the spatial locations of participants are distributed in three in three circular traffic zones with the same center but various radius. The inner most circle represent the highest traffic volume zone, while the outer most represents the lowest traffic volume zone. The distance between each passenger and driver are calculated by $D_{i,j} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}$. The gender of passengers and drivers and the review rating for drivers are uniformly distributed. Those above considerations lead to the following outline of the random instance generator.

Step 1: Determine the passenger over driver ratio R.

Step 2: After determine the population size of passenger I, we then create three traffic zones representing high, medium and low traffic volume representing different traffic conditions (i.e. higher traffic volume usually indicate more travel time and thus increasing the operation cost), with the corresponding radius $r1$, $r2$,

$r3$, respectively. 40% of the total requests were distributed within the high traffic volume zone, 40% were distributed within the medium traffic volume zone, 20% were assigned distributed the low traffic volume zone.

Step 3: Determine the population size of drivers (i.e., $J{=}I/R$) generate drivers' current location (i.e., longitude and latitude coordinates), and their distribution in the three circular traffic zones. The latter follows the same distribution as the traffic zone themself, i.e., 40% in high traffic volume zone,40% in medium traffic volume zone and 20% in low traffic volume zone.

Step 4: Generate the passengers destination coordinate information. For each request, the destination in a location with radius uniformly distributed between 2 and 20 miles from the center. The angle between each origin and destination is calculated as: angle(i) = max(0, angle(i-1)+$2\pi/I$), where I is the number of requests

3.4.2 Pareto solutions from Epsilon-method

In contrast to single-criterion optimization, there is no single global optimal solution in the field of multi-criterion optimization. Thus it is always necessary to determine a set of points that all fit a predetermined definition for an optimum. Without loss of generality, we assume that all the objective functions f_i for $i = \{1, \ldots ,k\}$ are for maximization problem. In general, the concept of optimality in multi-criterion optimization is referred as Pareto optimality or efficiency that is defined as:

Definition 1. *Pareto Optimality: A feasible solution $x^* \in X$ is Pareto optimal iff there does not exist another $x \in X$, such that $f(x) \geq f(x^*)$ and $f_i(x) > f_i(x^*)$ for at least one objective function.*

Often, solutions from some algorithms may not be Pareto optimal but may satisfy partial criteria, making them significant for particular applications. For

instance, the weakly Pareto optimality is defined as:

Definition 2. *Weakly Pareto Optimality: A feasible solution $x^* \in X$ is Weakly Pareto optimal iff there does not exist another $x \in X$, such that $f(x) > f(x^*)$ for all objective functions.*

In the literature, another commonly used method to solve multi-criterion optimization problem is known as ϵ-method. Given the following multi-criterion optimization problem:

$$\begin{aligned} \text{Max} \quad & f_1(x), f_2(x), ..., f_k(x) \\ \text{st.} \quad & x \in \mathbb{R}, \end{aligned}$$

where x is the decision variables, $f_k(x)$ are the k-th objective functions and $\mathbb{R}$ is the feasible region. In the ϵ-method, one of the objective functions is optimized by adding the rest objective functions as constraints,incorporating with e_k in the constraint as shown below:

$$\begin{aligned} \text{Max} \quad & f_1(x) \\ \text{st.} \quad & f_2(x) \geq e_2 \\ & f_3(x) \geq e_3 \\ & \cdots \\ & f_k(x) \geq e_k \\ & x \in \mathbb{R}. \end{aligned}$$

In order to apply the ϵ-method, one has to find the range for each individual objective function. One common approach is to construct the payoff table as shown in Table 2, where $x_k^* = \{\arg\max_x f_k(x) \mid x \in \mathbb{R} \ , \forall \ k \in K \ \}$.The best value can be easily attainable as the optimal as single optimization problem, while the worst value can be obtained as the minimum of the corresponding column.

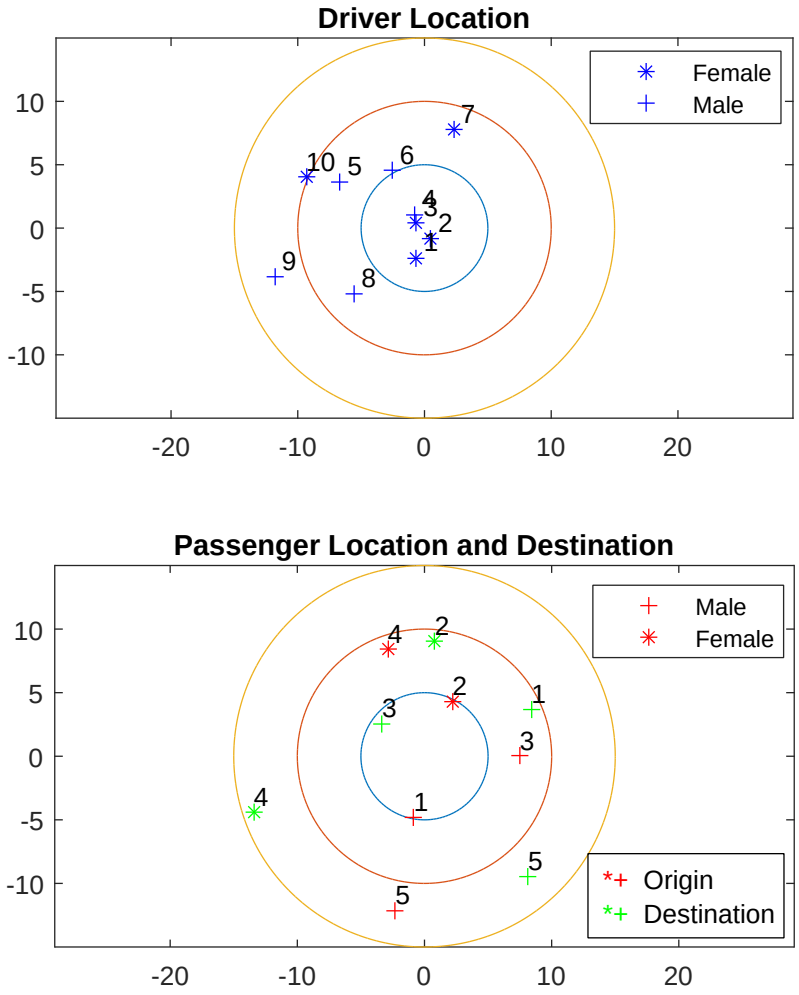

Figure 2. An illustrative example with 5 passengers & 10 drivers

3.4.3 An illustrative example

Let us consider the following example in Figure 2. The trip information for 5 passengers (2 female) and 10 drivers (5 female) are generated. For each potential pair, their profit matrix is given based on their location as shown in Table 3. Finally, the driver's review rating is presented in Table 4.

The final assignment is summarized in Table 5. First, notably all the female passenger's requests are received by female drivers, which indicates all female passengers will be served by female drivers only. This would address the female

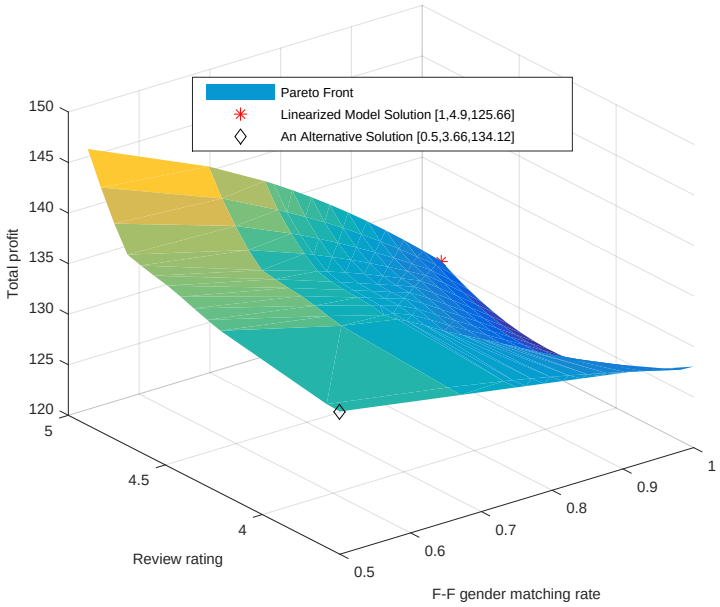

Figure 3. Pareto front of the illustrative example

passenger's safety concern when choose cloudsourcing transportation modes. Second, the average rating of selected driver from the optimal assignment is 4.9. Comparing to the non-optimized average rating at 3.9 (average rating from Table 4), this not only improve the customer experience but also highly motivate drivers to get highly review rating, in order to earn more business opportunity. Lastly, although profit is with the lowest priority comparing to gender matching rate and driver's review rating, the average profit per pair increase from \$10.9 (average profit from Table 3) to \$12.6.

We also use the ϵ-method to further validate the solution from the lexicographical approach. For each objective function f_i, we evenly divide the range to L equal grid intervals and obtain the corresponding value of ϵ as: $e_k^l = (f_k^* - \underline{f_k})\frac{(l-1)}{L} + \underline{f_k} \quad \forall l = \{1, 2, ..., L\}$, where f_i^* is the optimal value and $\underline{f_i}$ is the minimum of the corresponding column as demonstrated before (Table 2). Table 6 summarizes the 12 unique solutions obtained from ϵ-method and the Pareto front is

plotted in Figure 3. From Table 6, one concludes that the solution from linearized model solved by the lexicographical approach consist with ϵ-method. However it is also noted that all the constraints are active and hence all solutions correspond to pareto optimal solutions, from which the preferred solution may be selected. For example, the decision maker may compromise on lower average gender matching rate and average selected driver's rating to make more profit.

3.4.4 Numerical results

Next, we use five sets of random instances to test the proposed model's performance. In our experiments, each set consists of 20 randomly generated instances, where $|J|$ is set as 30 and $|I|$ varies between 21 and 45 corresponding to driver to passenger ratio between 0.7 and 1.5 as in Table 1. In addition, $\bar{P}$ is set as 2 and $\bar{D}$ is set as 3.

Table 7 and 8 summarize the overall performance across the five sets. Several observations can be made. First, as driver's size decreases from 45 to 21, the average gender matching rate decrease from 1 to 0.8805. This is obviously due to the decrease of available female drivers. Second, as the total amount of female drivers decreases, the average review rating of assigned driver also decrease from 4.29 to 3.12. Since in the objective function, the female passenger to female driver matching rate has the highest priority, the system has to compromise by selecting female drivers with lower review rating in order to satisfy female-passenger-to-female-driver matching rate. On the other hand, the number of drivers has marginal impact on the average profit per trip because it has the lowest priority. Furthermore, the driver size has very marginal impact on average profit per trip as it has the lowest priority. Furthermore, the improvement of our proposed model is remarkable in that the average review rating of assigned driver and average profit per trip has increased by 42%, 62%, 34%, 59%, 24%, 61%, 12%, 61%, 3%, 61%, respectively.

We also test how the maximum allowable requests per driver ($\bar{D}$), and

maximum allowable drivers per request ($\bar{P}$), affect the model solution. Using the same trip information generated from Cases 1, we re-run the multi-criterion model after changing those two parameters from 2 to 3 and 3 to 5, respectively, while all remaining all other constraints and parameters are kept the same. The results are displayed in Table 9 and 10. Overall, the performances are consistent with Case 1. As the number of drivers decreases, the average gender matching rate decreases, as well as the average review rating of selected driver. However, with the enlarged capacity, the decrease in gender matching is from 1 to 0.92, compared to from 1 to 0.88 in Case 1. Similarly, the decrease in average reviewing rating is from 4.39 to 3.32, compared to from 4.29 to 3.12 in Case 1. Again, the impact of reducing driver size on average profit is very limited and thus it is less sensitive comparing to the average gender matching rate and average review rating of assigned drivers.

TABLE 1

Parameter setting

Parameter	Value
I/J ratio	[1.5, 1.2, 1, 0.8, 0.7]
$\bar{P}$	[2,3]
$\bar{D}$	[3,5]

TABLE 2

Payoff table

	f_1	f_2	f_3	...	f_k
x_1^*	f_1^*	$f_2(x_1^*)$	$f_3(x_1^*)$	...	$f_k(x_1^*)$
x_2^*	$f_1(x_2^*)$	f_2^*	$f_3(x_2^*)$	...	$f_k(x_2^*)$
x_3^*	$f_1(x_3^*)$	$f_2(x_3^*)$	f_3^*	...	$f_k(x_3^*)$
...	...	...	...	...	...
x_k^*	$f_1(x_k^*)$	$f_2(x_k^*)$	$f_3(x_k^*)$	...	f_k^*

TABLE 3

Profit matrix

	$D1$	$D2$	$D3$	$D4$	$D5$	$D6$	$D7$	$D8$	$D9$	$D10$
$P1$	17.40	16.34	15.75	15.34	12.72	13.17	11.08	16.05	12.27	11.53
$P2$	3.04	4.15	4.49	4.74	2.04	4.55	5.27	0.03	2.34	0.46
$P3$	11.61	12.46	11.80	11.67	7.92	10.10	11.06	8.27	4.87	6.31
$P4$	18.30	19.02	19.91	20.33	21.25	22.57	21.79	16.59	15.81	20.21
$P5$	10.21	9.15	8.57	8.17	6.33	6.15	3.88	11.54	8.55	5.58

TABLE 4

Driver's review rating

	$D1$	$D2$	$D3$	$D4$	$D5$	$D6$	$D7$	$D8$	$D9$	$D10$
Rating	3	5	4	5	4	1	5	0	1	2

TABLE 5

Optimal assignment

	Passenger									
	1		2		3		4		5	
Driver	2	4	4	7	4	5	2	7	2	7

TABLE 6

Unique Pareto solutions from 20 grid intervals

Objective function	F_1	F_2	F_3
	0.5	3.6	134.12
	0.671	3.668	133.949
	0.75	4.216	131.921
	0.763	4.284	131.561
	0.789	4.353	131.142
	0.816	4.421	130.599
Unique pareto solution	0.842	4.489	129.953
	0.868	4.558	129.305
	0.895	4.626	128.656
	0.921	4.695	127.972
	0.947	4.763	127.267
	0.974	4.832	126.521
	1	4.9	125.66

TABLE 7

Female-female gender matching rate for case 1 ($\bar{P}$=2 $\bar{D}$=3)

Problem size	Female-female gender matching rate
$\|I\|$=30 $\|J\|$=45	1
$\|I\|$=30 $\|J\|$=36	1
$\|I\|$=30 $\|J\|$=30	0.9975
$\|I\|$=30 $\|J\|$=24	0.9615
$\|I\|$=30 $\|J\|$=21	0.8805

TABLE 8

Overall performance for case 1 ($\bar{P}$=2 $\bar{D}$=3)

	Proposed model		Non-optimized model		Improvement	
Problem size	Avg rating	Avg profit	Avg rating	Avg profit	Rating increament	Profit increament
$\|I\|$=30 $\|J\|$=45	4.29	44.63	3.02	27.52	42.05%	62.17%
$\|I\|$=30 $\|J\|$=36	4.03	44.21	3.01	27.65	33.89%	59.89%
$\|I\|$=30 $\|J\|$=30	3.67	44.73	2.94	27.77	24.83%	61.07%
$\|I\|$=30 $\|J\|$=24	3.34	44.01	2.98	27.31	12.08%	61.15%
$\|I\|$=30 $\|J\|$=21	3.12	44.28	3.02	27.45	3.31%	61.31%

TABLE 9

Female-female gender matching rate for case 2 ($\bar{P}$=3 $\bar{D}$=5)

Problem size	Female-female gender matching rate
$\|I\|$=30 $\|J\|$=45	1
$\|I\|$=30 $\|J\|$=36	1
$\|I\|$=30 $\|J\|$=30	1
$\|I\|$=30 $\|J\|$=24	0.9785
$\|I\|$=30 $\|J\|$=21	0.9200

TABLE 10

Overall performance for case 2($\bar{P}$=3 $\bar{D}$=5)

	Proposed model		Non optimized model		Improvement	
Problem size	Avg rating	Avg profit	Avg rating	Avg profit	Rating increament	Profit increament
$\|I\|$=30 $\|J\|$=45	4.39	42.62	3.02	27.52	45.36%	54.87%
$\|I\|$=30 $\|J\|$=36	4.16	42.11	3.01	27.65	38.21%	52.30%
$\|I\|$=30 $\|J\|$=30	3.83	42.66	2.94	27.77	30.27%	53.62%
$\|I\|$=30 $\|J\|$=24	3.53	41.72	2.98	27.31	18.46%	51.12%
$\|I\|$=30 $\|J\|$=21	3.32	42.19	3.02	27.45	9.93%	53.70%

CHAPTER 4

DRIVER'S PROSPECT MAXIMIZATION PROBLEM

4.1 Problem Statement

In this chapter, we develop a two-stage approach for an automatic process for optimizing operations for a large-scale collaborative ridesharing transportation platform. In the first stage, a centralized optimization-based approach is used to find the best driver-passenger matches through the best, possibly multiple, driver-passenger matches through the use of a multi-criterion optimization model. The assignment achieved in the first stage is a multiple-driver to multiple-passenger assignment, i.e., each driver can the assigned to multiple passengers and each passenger can be assigned to multiple drivers. In the second stage, a reverse auction process is used to model the decision making for drivers on to two decisions, i.e., the bidding price for each assigned passengers and selection from the winning bids. We use the Prospect Theory to model drivers desire to maximize his/her own prospect of profit making under uncertainty. Its also worth noting that the second stage is a decentralized model independently applied to all drivers.

The context we consider herein is to provide a decision framework as a recommendation tool to decide which task should be taken for those participating drivers working for an Uber-like platform. Prior to second-stage driver's prospect maximization module, the Uber-like platform in the first stage will provide an initial centralized task assignments that maximizes three criteria: average gender matching rate, average driver review rating and total system-wide profit. More specifically, in the first stage, the information including driver's current location, gender, review

rating, passengers current location, destination, gender, estimated traveling cost under current traffic condition and expected profit from this trip is given, a multi-criteria maximization problem is formulated to maximize: average gender matching rate, average driver's review rating and system-wide profit. However, it is likely that more than one driver is available and can be assigned to a passenger. This would cause them to compete in the bidding process in stage two. Next, in the second stage, prospect theory is employed to determine their bidding price for all passenger assigned to them in the first stage. As a result of the announcement of the bidding price, passenger-driver matching will be determined, assuming passengers always chose the lowest bidding price that received. However, note that drivers can decline a passenger if the prospect of serving the passenger is unattractive.

Without loss of generality, several assumptions are made as follows:

1. All drivers have their own knowledge for estimating the approximate traveling cost. Each driver will propose an exploratory charge to the passenger as a bidding price, the passenger will chose the lowest bid as the winning driver and thus any other drivers who placed a higher bidding price will lose this task. Possible scenarios faced by any driver is illustrated in Table 11. In this table, "passenger show" and "passenger not show" represent two ultimate outcomes, which is affected by driver's winning or losing a particular bid. The "revenue" and "loss" in the table are the corresponding expected net income, whereas P_j is the probability that the driver will win the bid for passenger j.

2. The procurement process is sealed and only allows each driver to submit their bidding price one time, once the driver places a bid, he/she is not able to revise his/her bidding price.

3. Each drivers bidding strategy is affected by the number of the competitors rather than competitors bidding price, that is, each driver is fully aware of how many potential competitors are reviewing the same request while he/she

is not able to know other drivers bidding price.

4. The drivers are allowed to submit multiple bids to different assigned passengers, but he/she is only able to serve one passenger. Thus he/she will always chose the passenger request with the highest prospect in his/her profit making.

4.2 Preliminaries on Prospect Theory

We have briefly reviewed literature on Prospect Theory(PT) in Chapter 2. In this section, we presents preliminaries of the Prospect Theory as it relates to this dissertation. Since its formulation by Kahneman and Tversky in 1979, prospect theory has emerged as a leading alternative to expected utility as a theory of decision making under risk. Prospect theory believes that individuals evaluate outcomes with respect to deviations from a reference point rather than with respect to net asset levels. Several distinct properties of the PT are listed below:

1. Outcomes are valued as gain or lose relative to a current reference point instead of final levels of wealth.

2. Under loss aversion scenario, the disutility of a loss can be greater than the utility of an equivalent gain.

3. The value of an outcome is weighted not by the probability of its occurrence, p, but by a weighted probability, $w(p)$.

Let $V(x_1, p_1 x_n, p_n)$ denote a prospect. The x_j, j=1, ..., n denotes the possible outcomes associated quantity of money, and p_j, j=1,,n denotes the corresponding probability for each outcome. The evaluation of original prospect theory is designed to deal with at most two outcomes, it is then calculated as:
$V(x_1, p_1, x_2, p_2) = u(x_1)w(p_1) + u(x_2)w(p_2)$, where $u(x)$ is the utility function

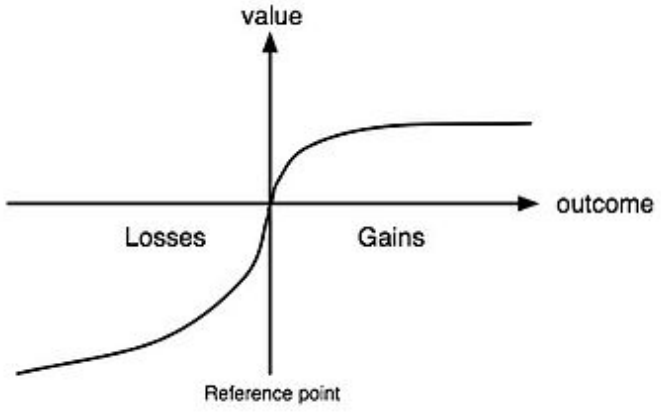

Figure 4. A Hypothetical Utility Function

associated with outcome x, describing the subject's valuation of money, and $w(p)$ is the weighted function associated with probability p that describes the subject's attitude towards probabilities. The utility function and weighted probability function used in our model are commonly seen in literature as:

Utility function:

$$u(x) = x^{\alpha} \quad (0 < \alpha < 1). \tag{29}$$

The utility function usually has three main characteristics. First, it is defined on deviations from a reference point, rather than on net asset position. Thus if the reference point shifts, the value function shifts accordingly. In this dissertation, we use $V(0) = 0$ as the reference point. Second, the utility function is generally concave for gains and convex for losses, reflecting risk aversion in the domain of gains and risk seeking in the domain of losses. Third, the loss curve can be steeper than gain curve by varying the coefficient α. A typical hypothetical utility function is illustrated in Figure 4.

Weighted probability function:

$$w(p) = e^{-(-lnp)^{\beta}} (\beta > 0) \tag{30}$$

The weighted probability function measures the impact of the probability of an event on the desirability of a prospect. Kahneman and Tversky [12] pointed out

that preferences of subjects can best be modeled by a weighting function that enhances small probabilities and reduces higher probabilities. Hence the weighting function is relatively sensitive to changes in probability near the end points 0 and 1, but is relatively insensitive to changes in probability in the middle region. Finally, the two function jointly determine the decision makers risk attitude.

4.3 Problem Formulation

Consider a set of driver I and a set of passenger J from tentative assignment generated by the multi-criteria optimization model in Stage 1, The decision variable b_{ij} represents the amount of bidding price from driver i to passenger j. In addition, let c_{ij} denote the distance based operational cost for driver i to delivery passenger j to destination and p_{ij} be the probability for driver i to win the request from passenger j. Further, α, β, γ are the coefficients in the utility utility function, weighted probability function and driver's attitude towards risk, respectively. We can then formulate the driver's prospect problem as:

$$Maximize \qquad (b_{i,j} - c_{ij})^{\alpha} e^{-(-lnp_{i,j})^{\beta}} - \lambda (c_{ij})^{\alpha} e^{-(-ln(1-p_{i,j}))^{\beta}} \tag{31}$$

$$s.t. \qquad \underline{B} \leq b_{ij} \leq \overline{B} \tag{32}$$

The objective function aims to maximize the prospect for driver i to pick up passenger j, while the drivers bid price is limited by the lower and upper bounds. The winning probability $p_{i,j}$ is determined as follows. Suppose n competitive drivers bidding for the same passenger request are identical and independently distributed. These drivers draw their bidding price from some random distribution $\Phi\{f(x),F(x)\}$, (e.g., uniform distribution). Then, the probability for driver i to win a request is the probability that no other bidding price is lower than driver i's, which equals $[1 - F(b_i)]^{n-1}$.

Theorem 3. *Let* $y = (b_{i,j} - c_{ij})^{\alpha} e^{-(-lnp_{i,j})^{\beta}} - \lambda (c_{ij})^{\alpha} e^{-(-ln(1-p_{i,j}))^{\beta}}$. *If* y *is strictly*

concave, then the driver's prospect maximization model has a unique global solution b^*.

Proof. Clearly, the feasible region $\underline{B} \leq b_{ij} \leq \overline{B}$ is bounded and convex. Thus, any local maximum is the global maximum and the uniqueness of the global maximum follows immediately from the strict convexity of the objective function. □

Theorem 4. *if* y *is twice differentiable on the feasible region* $\underline{B} \leq b_{ij} \leq \overline{B}$ *and* $\frac{d^2y}{db_{i,j}^2}$ *strictly less than 0 , then the driver's prospect maximization problem is concave.*

Proof. Let $A = e^{-(-lnp_{i,j})^\beta}$, and $B = e^{-(-ln(1-p_{i,j}))^\beta}$. The second order derivative of the objective can be written as:

$$
\begin{aligned}
\frac{d^2y}{db_{i,j}^2} = & \underbrace{\alpha(\alpha-1)(b_{i,j}-c_{ij})^{\alpha-2}A}_{<0} \\
& + \underbrace{\alpha(b_{i,j}-c_{ij})^{\alpha-1}A(\beta(-ln(\frac{\overline{B}-b_{i,j}}{\overline{B}-\underline{B}})^{\beta-1})\frac{-1}{\overline{B}-b_{i,j}})}_{<0} \\
& + \underbrace{(b_{i,j}-c_{ij})^{\alpha}A(\beta(-ln(\frac{\overline{B}-b_{i,j}}{\overline{B}-\underline{B}})^{\beta-1})\frac{-1}{(\overline{B}-b_{i,j})^2})}_{<0} \\
& + \underbrace{(b_{i,j}-c_{ij})^{\alpha}A(\beta(\beta-1)(-ln(\frac{\overline{B}-b_{i,j}}{\overline{B}-\underline{B}})^{\beta-2})\frac{1}{(\overline{B}-b_{i,j})^2})}_{<0} \\
& + \underbrace{\lambda(c_{ij})^{\alpha}B(\beta(-ln(\frac{b_{i,j}-\underline{B}}{\overline{B}-\underline{B}})^{\beta-1})\frac{-1}{(\overline{B}-b_{i,j})^2})}_{<0} \\
& + \underbrace{\lambda(c_{ij})^{\alpha}B(\beta(\beta-1)(-ln(\frac{b_{i,j}-\underline{B}}{\overline{B}-\underline{B}})^{\beta-2})\frac{1}{(b_{i,j}-\underline{B})^2})}_{<0} \\
& + \underbrace{\lambda(c_{ij})^{\alpha}B(\beta(-ln(\frac{b_{i,j}-\underline{B}}{\overline{B}-\underline{B}})^{\beta-1})\frac{-1}{(B_{i,j}-\underline{B})^2})}_{<0}.
\end{aligned}
$$

From the above derivation, the objective function y is twice differentiable at $b_{i,j} \in \{Lb, Ub\}$ and $\frac{d^2f}{db_{i,j}^2} < 0$, thus y is concave and decreasing function. □

In order to characterize the optimal solution b^*, we introduces the Lagrangian multipliers $\mu_{i,j}$ and $\phi_{i,j}$ for lower and upper bound constraints respectively. Consequently, the driver prospect maximization problem can be rewritten as the following complementarity problem:

$$\alpha(b_{i,j} - b_{ij})^{\alpha-1} e^{-(-lnp_{i,j})^{\beta}} - (b_{i,j} - c_{ij})^{\alpha} e^{-(-lnp_{i,j})^{\beta}} \beta(-(-lnp_{i,j})^{\beta-1}) \frac{1}{Ub - b_{i,j}}$$
$$-\lambda(c_{ij})^{\alpha} e^{-(-ln(1-p_{i,j}))^{\beta}} \beta(-(-ln(1-p_{i,j}))^{\beta-1}) \frac{1}{b_{i,j} - Lb} = 0 \quad (33)$$

$$\mu_{i,j}(b_{i,j} - \underline{B}) = 0 \quad (34)$$

$$\phi_{i,j}(\overline{B} - b_{i,j}) = 0 \quad (35)$$

$$\mu_{i,j} \geq 0 \quad (36)$$

$$\phi_{i,j} \geq 0 \quad (37)$$

$$\underline{B} \leq b_{ij} \leq \overline{B} \quad (38)$$

The above complementarity problem in fact defines the Karush-Kuhn-Tucker(KKT) optimality conditions for the optimal solution b^* and associated Lagrangian multipliers. Note that formulating the KKT conditions allows for an alternative efficient solution by the nonlinear program software GAMS [80] used in this study.

4.4 Computational Results

4.4.1 An illustrative example

Recall the illustrative example in Chapter 4 as shown in Figure 2. After solving the multi-criterion optimization model in Stage 1, the optimal assignment is then immediately used in the driver's prospect maximization problem. This two-stage procedure is summarized as follows:

Input: P - tentative many-to-many assignment :
Output: F - final driver-passenger assignment :
for $i \in P$ **do**:
    Solve max z_j $|\underline{B} \leq b_{ij} \leq \overline{B}$
end for

With the tentative driver-passenger assignment from Stage 1 as displayed (again) in Table 12, Table 13 shows the step-by-step auction process until the last passenger's request is fulfilled. Particularly in this table, The first four columns "Passenger", "Driver", "Origin-Destination" and "Distance between P&D" represent the spatial relationship between each tentative assignment pair. Columns "Lower bound" and "Upper bound" represent the lowest and highest bidding price that a driver can place for a certain passenger. $\underline{B}_{ij} = \underline{R} \cdot C_{ij}$ and $\overline{B}_{ij} = \overline{R} \cdot C_{ij}$, where $\underline{R} \sim U[0.8, 11]$ and $\overline{R} \sim U[1.1, 1.5]$ are lower and upper bounds for fare rate (\$/mile). Both $\underline{B}_{ij}$ and $\overline{B}_{ij}$ are rounded to nearly integers. Finally, last five columns "# of bidders", "# of biddings", "Final bidding", "Announced winning driver" and "Prospect" report the total number of bidders competing for the same passenger's request, the total number of request each bidder currently bidding on, each bidder's bidding price on each assigned request, the final announced winning driver and the corresponding prospect for each driver to take that request, respectively.

Note that in this illustrative example, Passengers 1 and 4 are picked up by driver 4 and 7. Note that Driver 4 wins both Passenger 1 (bid 55.85), Passenger 2 (bid 39.62) and Passenger 3 (bid 58.68) by offering the lowest bid (when compared to bids from Driver 2 for Passenger 1 (bid 57.82) and bids from Driver 7 for Passenger 2 (bid 41.00) and bids from Driver 5 for Passenger 3 (bid 60.55). Once Driver 4 secures both bids, he decides to pick up passenger 1 due to its highest prospect of 1.17, compared to the prospect of 0.96 for picking up Passenger 2 and prospect of 1.11 for picking up Passenger 3. Similar situation occurs for Driver 7,

whose final decision is to pick up Passenger 4 due to higher prospect in return, when winning both Passenger 4 and 5 requests.

TABLE 11

Driver's choice

Driver's potential choice		Passenger show	Passenger not show
Pick passenger 1	Profit	Revenue(+)	Loss(-)
	Prob	P_1	1 - P_1(Lose)
Pick passenger 2	Profit	Revenue(+)	Loss(-)
	Prob	P_2	1 - P_2
Pick passenger j	Profit	Revenue(+)	Loss(-)
	Prob	P_j	1 - P_j
No passenger picked	Profit		0
	Prob		1

TABLE 12

Optimal assignment

Passenger	Driver
1	2
	4
2	4
	7
3	4
	5
4	2
	7
5	2
	7

TABLE 13

Step-by-step auction process in stage 2 for illustrative example

Bidding starts with tentative assignment from stage 1										
Passenger	Driver	Origin-Destination	Distance between D&P	Lower bound	Upper bound	# of bidders	# of bidding	Final bidding	Announced winning driver	Prospect
1	2	54.88	3.41	52	69	2	3	57.82		1.07
1	4	54.88	1.20	52	69	2	3	55.85	***4***	***1.17***
2	4	38.09	1.58	36	48	2	3	39.62	4	0.96
2	7	38.09	3.21	36	48	2	3	41.00		0.87
3	4	56.33	2.69	54	70	2	3	58.68	4	1.11
3	5	56.33	4.88	54	70	2	1	60.55		1.01
4	2	51.39	4.52	49	64	2	3	55.26		0.97
4	7	51.39	1.24	49	64	2	3	52.42	***7***	***1.13***
5	2	14.77	4.80	16	20	2	3	18.62		0.31
5	7	14.77	3.27	16	20	2	3	17.47	7	0.47
Remove passenger 1&4 and driver 4&7.										
2	2	38.09	4.64	36	48	2	3	42.15		0.79
2	3	38.09	2.69	36	48	2	2	40.56	***3***	***0.90***
3	2	56.33	4.12	54	70	2	3	59.91	***2***	***1.05***
3	5	56.33	4.88	54	70	2	1	60.55		1.01
5	2	14.77	4.80	16	20	2	3	18.61		0.31
5	3	14.77	3.08	16	20	2	2	17.32	3	0.49
Remove passenger 2&3, driver 3&2, bidding ends until passenger 5 is served.										
5	5	14.77	2.79	16	20	2	1	17.09	***5***	***0.52***
5	1	14.77	3.07	16	20	2	1	17.31		0.49

4.4.2 Numerical results

In this section, we use the three sets of random instances generated in Section 3.4 to further test the proposed driver prospect maximization problem. Recall that each set consists of 20 randomly generated instances. Those passenger sizes are fixed at 30 while the drive sizes are varied from 21, 30 and 45 respectively, according to different driver/passenger ratio from 0.7,1 and 1.5; and the maximum allowed request per driver and the maximum allowed driver per passenger is setting to 5. Table 14 through Table 16 shows shows the aggregate measures for winning pairs from Set 1 through Set 3. First two columns represent passenger and driver indices, Columns Bid and Prospect are the drivers bidding price for each passenger and the corresponding prospect based on this bidding price. Columns WP to WLP are the probabilities of winning, losing, weighted probability of winning and weighted probability of losing, respectively. The last three columns display that the expected profit, the potential loss and the percentage of passengers' requests being fulfilled, respectively. Several observations can be made. First, as the number of drivers increases from 21 to 45, the percentage of requests being fulfilled increases from 16% to 27% Second, across all the five sets, the high winning probability (most of them are above 98%) indicates all the drivers are rather risk-averse when placing their bids, this could be explained as the drivers try to place their bids cautiously in order to guarantee higher winning probability and to maximize the corresponding prospect.

TABLE 14

Winning driver and passenger pair for set 1

Passenger	Driver	Bid	Prospect	WP	LP	WWP	WLP	Expected profit	Expected loss	Coverage percent
1	15	53.92	1.90	99.34%	0.66%	93.94%	4.95%	9.18	0.82	
15	2	54.89	1.06	98.54%	1.46%	92.99%	5.66%	9.53	0.97	13%
26	21	57.83	1.10	99.07%	0.93%	93.63%	5.19%	9.81	1.20	
30	5	61.39	0.98	96.80%	3.20%	90.99%	7.16%	10.23	3.34	

TABLE 15

Winning driver and passenger pair for set 2

Passenger	Driver	Bid	Prospect	WP	LP	WWP	WLP	Expected profit	Expected loss	Coverage percent
1	15	53.92	1.90	99.34%	0.66%	93.94%	4.95%	9.18	0.82	
4	11	50.61	1.05	98.95%	1.05%	93.48%	5.29%	8.65	0.84	
6	13	50.38	0.97	97.65%	2.35%	91.95%	6.43%	8.45	2.09	20%
15	2	54.89	1.06	98.54%	1.46%	92.99%	5.66%	9.53	0.97	
26	21	57.83	1.10	99.07%	0.93%	93.63%	5.19%	9.81	1.20	
30	5	61.39	0.98	96.80%	3.20%	90.99%	7.16%	10.23	3.34	

TABLE 16

Winning driver and passenger pair for set 3

Passenger	Driver	Bid	Prospect	WP	LP	WWP	WLP	Expected profit	Expected loss	Coverage percent
1	15	53.92	1.09	99.34%	0.66%	93.94%	4.95%	9.19	0.82	
4	11	50.61	1.05	98.95%	1.05%	93.48%	5.29%	8.65	0.85	
6	13	50.38	0.97	97.65%	2.35%	91.95%	6.43%	8.46	2.10	
11	3	53.05	1.06	98.96%	1.04%	93.49%	5.29%	9.04	1.25	27%
15	2	54.89	1.06	98.54%	1.46%	92.99%	5.66%	9.54	0.98	
19	16	56.85	1.08	98.82%	1.18%	93.32%	5.42%	9.65	1.37	
26	21	57.82	1.10	99.07%	0.93%	93.63%	5.19%	9.81	1.20	
30	5	61.33	0.98	96.80%	3.20%	90.99%	7.16%	10.24	3.34	

We also test how α and λ, the utility coefficient and the coefficient of driver's attitude towards risk, affects the drivers bidding behavior. Using the same trip information generated from set 1, we re-run the driver's prospect maximization model under five different scenarios by varying α in the range of 0.25 and 0.5, λ in the range of 0.75 and 1.5, while β is kept the same. The parameters setting for extensive sensitivity analysis are given in Table 17.

Table 18 summarizes the results for set 1 under scenarios 1 to 5. Among which scenario 2 and 3 varies utility function coefficient α from 1/2 to 1/3 and scenario 4 and 5 varies negative outcome coefficient from 0.75 to 1.5, while other parameters are fixed. First of all, the optimal passenger-driver pair is consistent with scenario 1. This shows that all the pairs remains the same although we vary α from 1/4 to 1/2 as well as λ from 0.75 to 1.5. Secondly, comparing scenario 1, 2 and 3, when α varies from 1/4 to 1/2, the bidding price from each driver moderately increases by 0.46% (scenario 2 comparing to scenario 1), 2.06% (scenario 3 comparing to scenario 1) on average. Consequently, the prospect for each driver increase by 9.99% (scenario 2 comparing to scenario 1), 34.33% (scenario 3 comparing to scenario 1) on average whereas the winning probability for each driver decreased by 0.57% (scenario 2 comparing to scenario 1), 2.46% (scenario 3 comparing to scenario 1) on average. This is because the increment for utility function is increasing exponentially while the utility function coefficient increase from 1/3 to 1/2. With this drastic increment, the drivers are able to place more aggressive bidding price and thus their corresponding winning probability is decreasing as shown in Figure 5

Similarly, Figure 6 depicts the impact by varying the negative outcome coefficient λ from 0.75 to 1.5, while the utility coefficient α remains fixed as 1/3. From the figure, we can see that the amplification of prospect for each driver is marginal comparing to the amplification in Figure 5. In contrast, the negative outcome coefficient λ is less sensitive comparing to the utility function α in terms of

outcomes, thus there is less room to reduce the bidding price when increase from 0.75 to 1.5. Hence the prospect curves for each driver are less steep than in Figure 5. On the other hand, with all the defensively placed bidding price, the overall winning probability slightly increased from 96% to 99%.

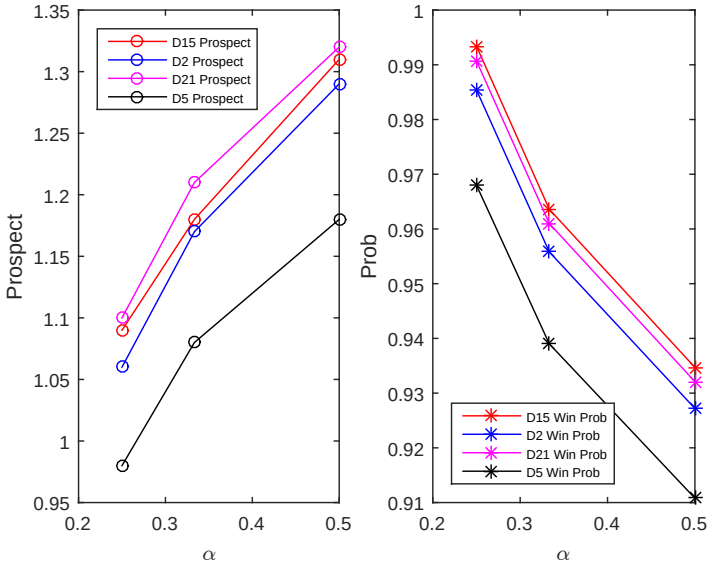

Figure 5. α varies from 0.25 to 0.5

TABLE 17

Prospect function parameter setting

	a	b	c
Scenario 1	1/4		1
Scenario 2	1/3		
Scenario 3	1/2	1.2	
Scenario 4	1/3		0.75
Scenario 5	1/3		1.5

TABLE 18

Winning driver and passenger pair for set 1 under Scenario 1 to 5

	Passenger	Driver	Bid	Prospect	WP	LP	WWP	WLP	Expected profit	Expected loss
s1	1	15	53.92	1.09	99.34%	0.66%	93.94%	4.95%	9.18	0.82
	15	2	54.89	1.06	98.54%	1.46%	92.99%	5.66%	9.53	0.97
	26	21	57.83	1.1	99.07%	0.93%	93.63%	5.19%	9.81	1.2
	30	5	61.39	0.98	96.80%	3.20%	90.99%	7.16%	10.23	3.34
s2	1	15	55.75	1.18	96.36%	3.64%	91.12%	7.74%	9.02	1.51
	15	2	57.79	1.17	95.58%	4.42%	90.20%	8.21%	9.31	1.74
	26	21	58.51	1.21	96.10%	3.90%	90.82%	7.78%	9.75	3.47
	30	5	46.39	1.08	93.90%	6.10%	88.26%	9.73%	8.74	2.76
s3	1	15	59.56	1.31	93.47%	6.53%	88.39%	9.28%	8.64	3.23
	15	2	63.97	1.27	92.72%	7.28%	87.49%	9.94%	8.80	3.69
	26	21	60.36	1.32	93.21%	6.79%	88.10%	9.33%	9.54	12.47
	28	5	46.91	1.18	91.08%	8.92%	85.61%	10.17%	8.43	2.56
s4	1	15	57.06	1.26	95.79%	4.21%	90.71%	7.59%	8.90	2.06
	15	2	59.54	1.22	95.04%	4.96%	89.82%	7.92%	9.17	2.26
	26	21	59.00	1.27	95.54%	4.46%	90.42%	7.73%	9.69	5.53
	28	5	46.53	1.13	93.40%	6.60%	87.94%	9.28%	8.27	2.37
s5	1	15	53.81	1.22	98.12%	1.88%	93.03%	6.34%	9.16	0.8
	15	2	54.44	1.18	97.36%	2.64%	92.14%	5.47%	9.49	0.94
	26	21	57.69	1.23	97.86%	2.14%	92.74%	4.98%	9.77	1.12
	30	5	61.42	1.09	95.73%	4.27%	90.26%	8.54%	10.17	3.23

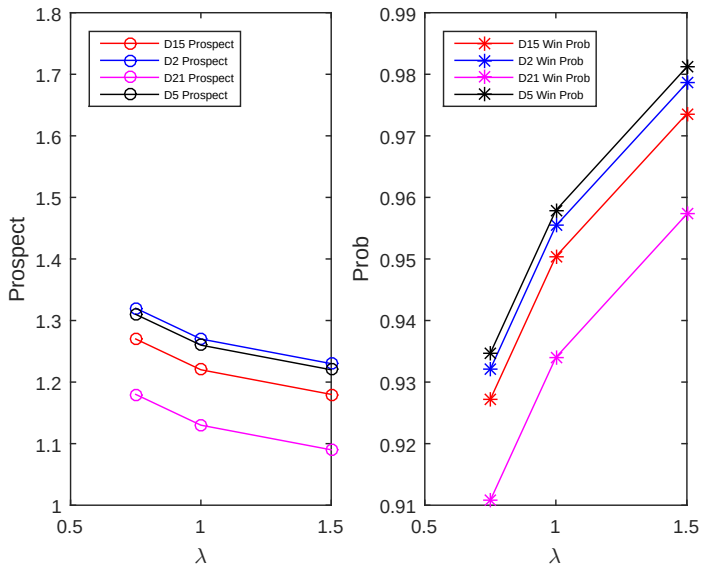

Figure 6. λ varies from 0.75 to 1.5

CHAPTER 5

COMPUTATIONAL EXPERIMENTS WITH NEW YORK CITY TAXI DATA

5.1 Problem Statement

To further test our proposed collaborative transportation framework, we analyze the taxi yellow cab data collected from New York city in 2017 with information from both demand and supply perspectives. We collected relevant data to perform the proposed analysis on the taxi supply and demand forecasting models within a given urban area. Those data mainly includes two part. First, the complete information for all the trips collected in NYC metro area in April 2014. Second, the trip data are supplemented with demographic and employment data from US Census Bureau, which includes key characteristics of the locations. The data description and preparation details are given in details in the following sections.

5.1.1 Data description

The first data set we used contains records of 14 million taxi trip information, i.e., time stamp, location, travel time, distance and fare paid, between April 1 and April 30, 2014. Specifically, the dataset includes 18 columns as follows:

1. Vendor ID;
2. Pickup latitude and longitude, ate, and time;
3. Drop-off latitude and longitude, date, and time;

4. Distance Traveled from pickup to drop off;

5. Number of passengers;

6. Total Fare paid, including breakdown by fare, tolls, and tips;

7. Method of payment (e.g., cash, credit card).

These data are collected in the New York City Taxi & Limousine Commission (TLC) using the GPS and meter devices that are installed in all licensed taxi. In addition to the taxi trip information, the social and economical characteristics of the areas where taxi trips start and end that are likely to have an effect as well. In general, including size and demographic characteristics of the population and employment associated with a traffic analysis zone (TAZ) can improve the explanatory power for predicting trips for a given TAZ. Most of these data are collected by government entities such as the United States Census Bureau every ten years. The population data is often categorized by race, ethnicity age, etc. Information is also available on gender, housing cost, home value, poverty, employment status, and marital status. The employment data is also important in this study because the number of jobs is a good indicator of peoples economic and social activities. Where there are jobs, there is a need to use transportation to go to work and participate in economic activities. Therefore, in this study, we consider population, household mean income and number of employed population obtained from *American FactFinder.com*, as three input/independent variables in predicting taxi demand. Since most of those characteristics are aggregated at the level of census tracts, in this dissertation we aggregate the trips at the census tract level in our data preparation.

5.1.2 Data preparation

The raw data undergoes cleaning and pre-processing for subsequent use in developing demand forecasting models. First, all the records with pickups location outside NYC metro areas is removed to avoid price variability due to different fare rate. Second, duplicate records are removed as well as records with missing information, considering the number of records with missing information is very marginal comparing to the whole data set. Lastly, we remove all records deemed invalid using four criteria. First, either one of total fare amount, travel distance or travel time is zero. Second, the ratio between total fare, travel time and travel distance is unreasonably large or small. Third, if the pickup or drop off location contains invalid or null longitude and latitude information. Fourth, the fare amount (excluded tip) to distance ratio is calculated, those fare to distance ratio greater than 20 dollar per mile are considered as outlier and thus should be removed. The detailed breakdown for each step is given in Table 19.

Finally, the data set of demographics contains the total population, average house hold income and number of employed across the 2167 census tracts in 2010 were obtained from *American FactFinder.com*. Ultimately, 116 census tract with insufficient either population, average house hold income or employment information were removed from future analysis.

5.2 Regression Models

5.2.1 Price estimation

In general, the price of a ride can be modeled as a function of distance, with a constant upfront charge. In this section, we attempt to examine and understand the collected data to be used as input the proposed two-stage Uber-like trip distribution models. To better understand the problem, an exploratory analysis is performed on the processed trip data. Figure 7 demonstrates the overall frequency

of pickups(hourly) distribution for the one month data. Here the rush hour is identified as 7 pm - midnight. According to the identified rush hours, Figure 8 and Figure 9 plot the price against the trip distance during rush hours and non-rush hours. As expected, both figures show a strong linear correlation between price and trip distance. To avoid biased regression results between rush hours and non-rush hours, after pre-processing, the trip data is divided into two sets: rush hour set and non-rush hour set. Since we only consider distance as the explanatory variable, univariate linear regression is used to predict the relationship between price and distance in the form: $y(price) = \alpha + \beta x(distance)$

To avoid overly trained model, we split each data set into two sets, training set (70%) and test set (30%), for both rush hours data set and non-rush hours data set. The size of each data set and univariate linear regression results are given in Table 20. For the rush hours and non-rush hours models, there are more than 8 millions observations and more than 5 millions observations were used in the model development, respectively. From the table, both linear regressions yield rather high R squared values at 0.9044 and 0.9055, with mean square error of 3.53 and 3.85. respectively. Further, the regression models suggest that the difference between rush hour and non rush hour for taxi price is very marginal.

Lastly, the linear regression diagnostics is checked by plotting the residual error against the fitted value. From Figure 10 and Figure 11, the residuals are fairly randomly distributed across the center line, practically confirming normality and independence.

5.2.2 Demand forecasting

In this section, the demand forecasting models are developed to identify the taxi demands within certain census tracts and hours. As discussed previously, taxi demand is related to the demographic characteristics of certain areas and the time of the day. Thus three theoretically important variables considered are the

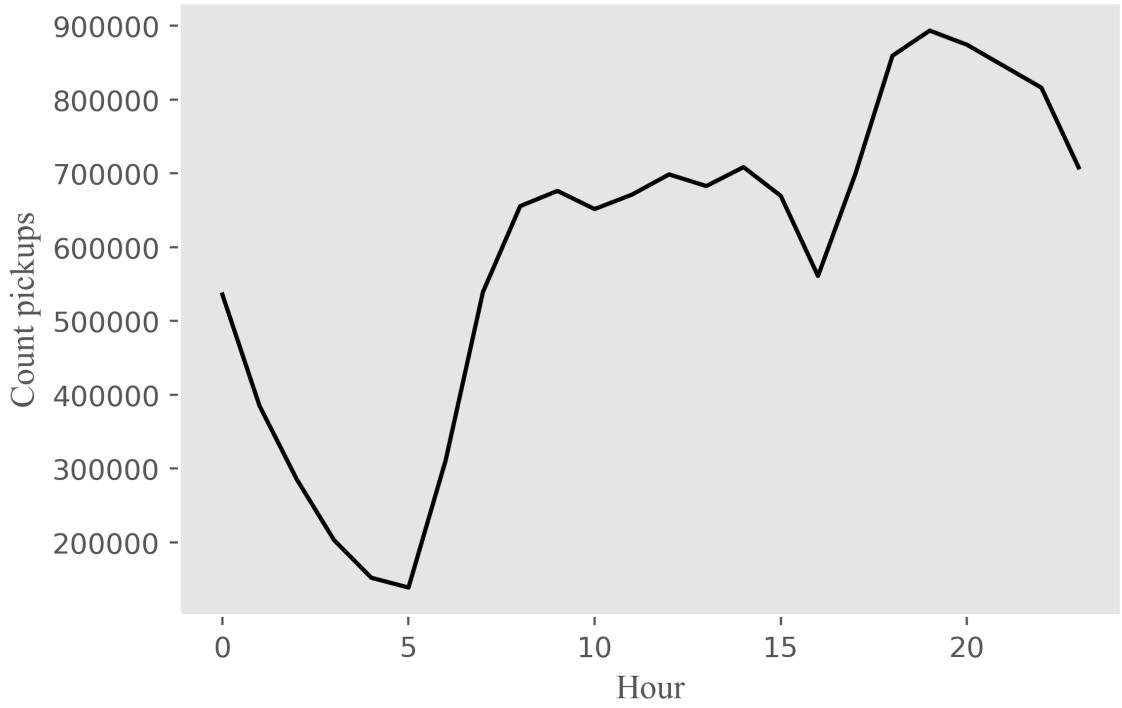

Figure 7. Hourly distribution of pickups

population, average household income and number of employed population. Also the dependent variable, taxi demand is aggregated into rush hours identified from the previous section. Finally, the transportation analysis zones in this dissertation are census tracts, thus all of the data is grouped by census tract so that the dependent and independent variables are aggregated at the same spatial resolution.

On the other hand, linear models have been widely used in demand forecasting [74], although literature suggests linear regression is not appropriate [xx]. In order to predict the cumulative count of random events, a generalized linear model is introduced by linking and the count of random events and independent variables through a log function:

$$\log(Y) = \beta_0 + \sum_i \beta_i x_i, \tag{39}$$

where Y is the total number of taxi demand generated in a census tract and x_i are the independent variables, β_0 is the intercept while β_i are the coefficients

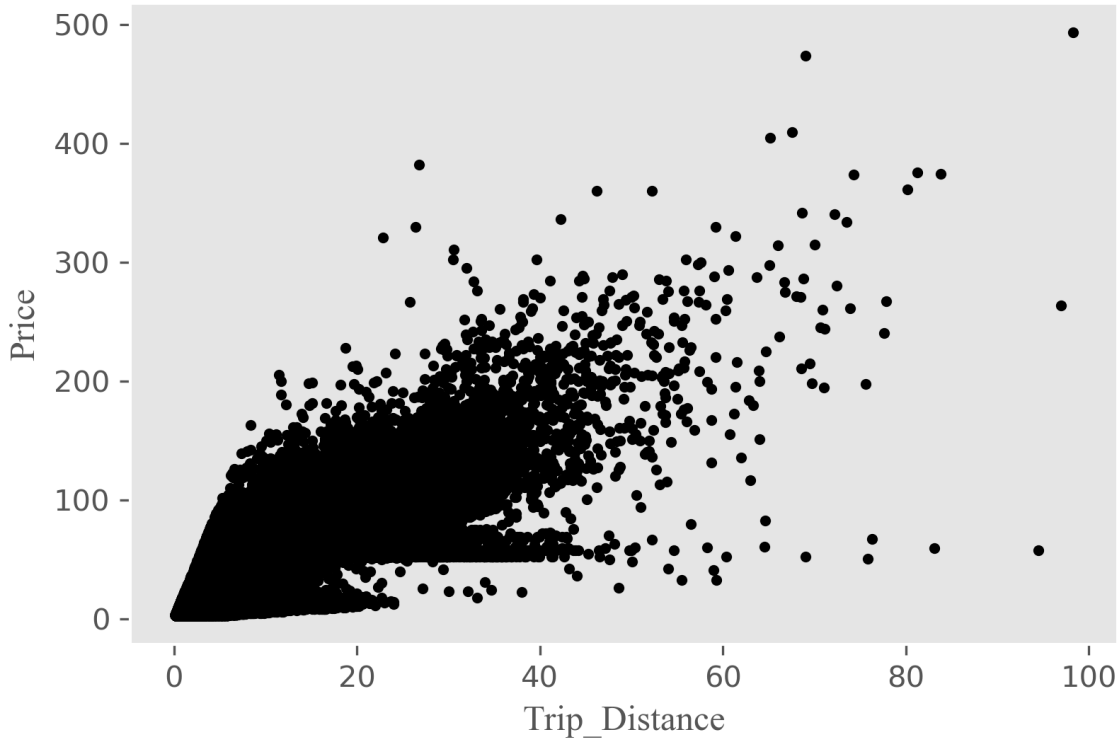

Figure 8. Price vs. trip distance during rush hours

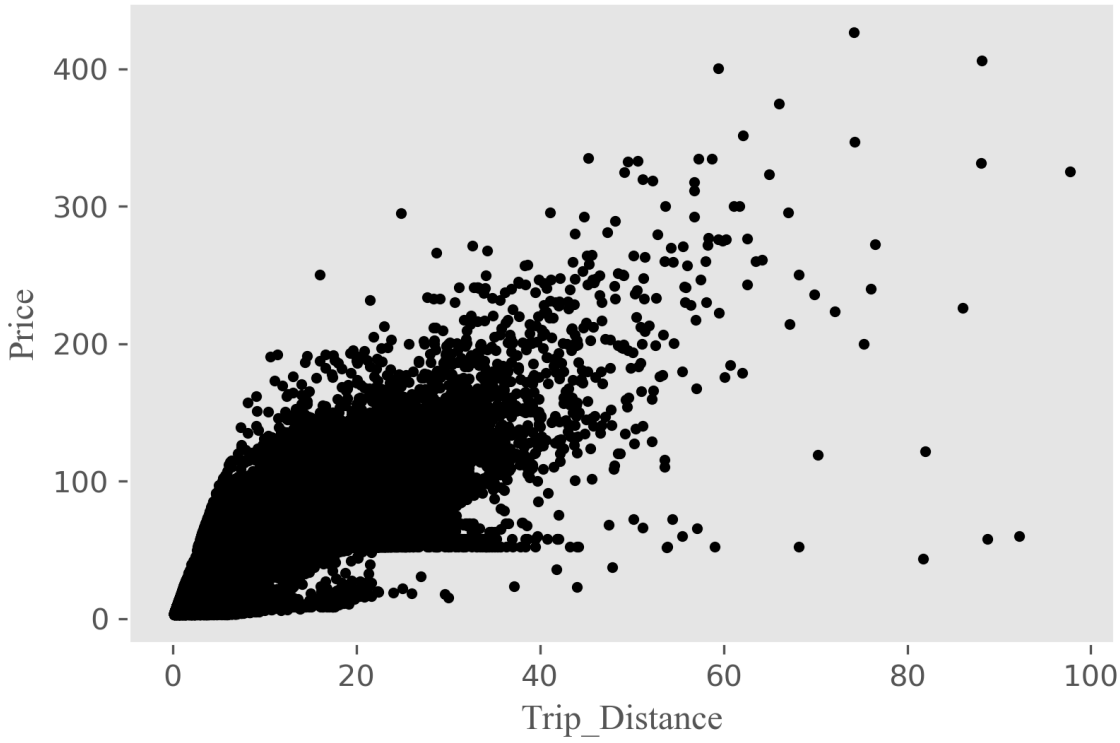

Figure 9. Price vs. trip distance during non-rush hours

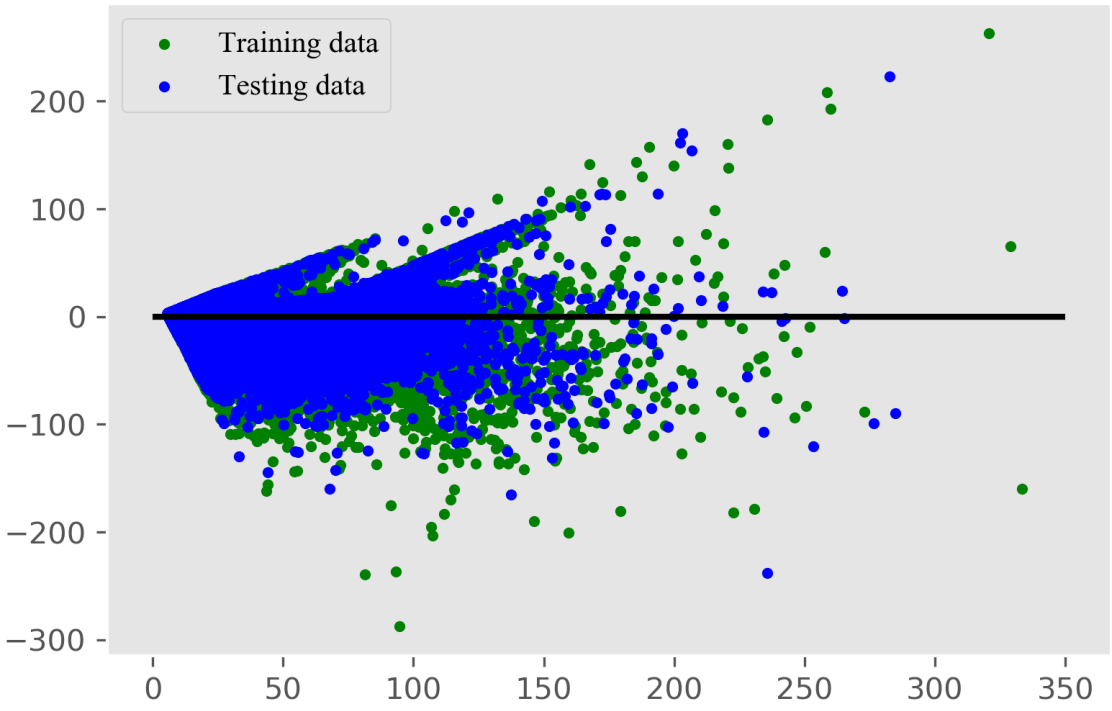

Figure 10. Residual vs. fitted value for rush hour model

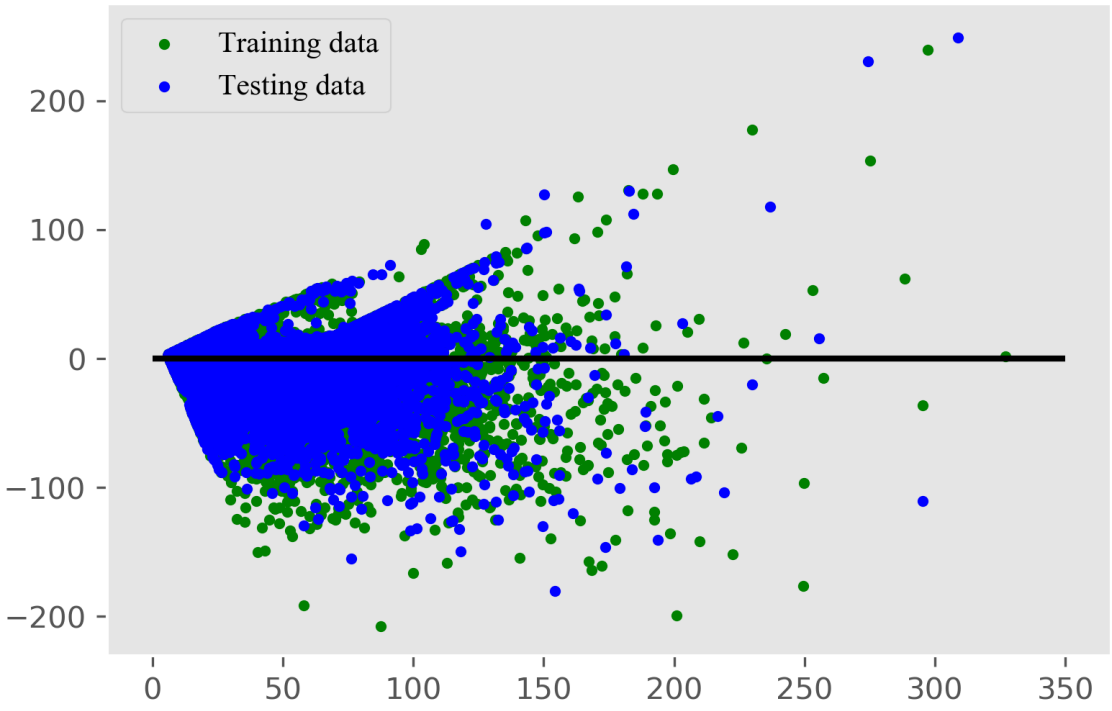

Figure 11. Residual vs. fitted value for non-rush hour model

corresponding to x_i. There are two commonly used regression models for counting random events, i.e., Poisson and negative binomial regressions. The Poisson regression assumes that the random events follow a Poisson distribution and the mean and variance of the random variable Y is given:

$$\mathrm{E}(Y_i) = \mu_i, \tag{40}$$

$$\mathrm{Var}(Y_i) = \theta\mu_i, \tag{41}$$

where θ is the dispersion parameter. The model is Poisson model when $\theta = 1$. On the other hand, the negative binomial model is characterized by a quadratic relationship between the variance and mean of the dependent variable [81]. For the negative binomial model, the mean of Y is the same as above but the variance is:

$$\mathrm{Var}(Y_i) = \theta\mu_i^2 + \mu_i, \tag{42}$$

In order to choose the distribution that most appropriately represents the count of random pickups, the data has been processed as follows:

1. By plotting the number of pickups on each day of the week, it clearly shows that weekdays have relatively more pickups than weekends in Figure 12. Thus all the trip records during weekends are removed from further regression model development;

2. All the remaining data are aggregation into hourly interval and then divided into 10 subsets sorted by count of pickups, the top 10% of the highest pickup counts are selected to develop the regression model, Finally, 328 samples were used to train the regression model. A sample subset of input data is displayed in Table 21.

The Poisson regression and negative binomial regression results are summarized in Table 22 and 23. First, both models conclude that population, mean

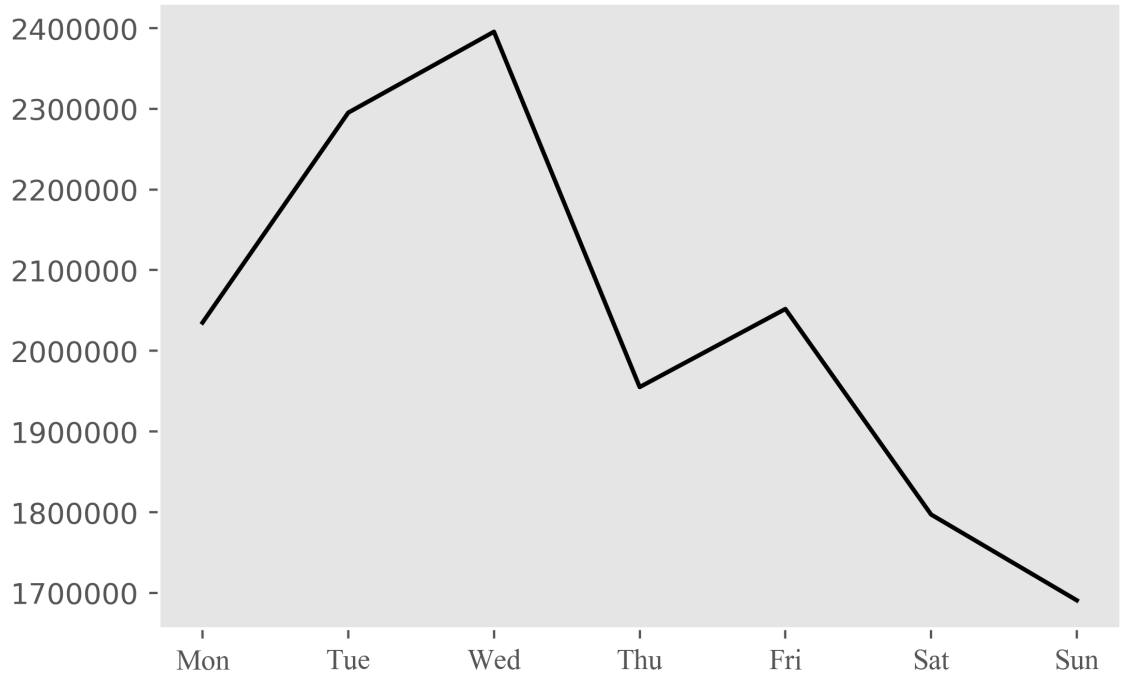

Figure 12. Pickups distribution by day of week

household income, number of employed people and time of the day are the significant variables. As expected, the population, mean household income and number of employed people have a positive correlation to the demand of taxi. In other words, the areas with higher population density, higher average income and more job opportunity tend to have higher taxi demand, comparing to those areas with lower demographic or economical attributes. Second, in addition to the demographic variables, time of the day tends to be critical in traffic planning as the results indicates: the taxi demands during rush hours (8:00 am to 9:00 am, 5:00 pm to 10:00 pm) are generated more than non rush hours.

In order to determine the most appropriate regression model, three performance metrics were compared between Poisson regression and negative binomial regression models. They are:

1. The Akaike Information Criterion (AIC). It is a measure of the relative quality of statistical models by trading off the complexity and goodness of fit. A

smaller AIC value represents a better model. The measure is then used to ensure that the model is not overfitted to the data.

2. The sum of squared deviance residuals.It is a measure of model fit for count regressions. The sum of model deviances is calculated as: $G^2 = 2\sum_i y_i \ln\frac{y_i}{\mu_i}$, where y_i is the observed value and μ_i is the prediction. In general, a smaller G^2 indicates a better model fitting.

3. Log-Likelihood is a probability that is used to describe a function of a parameter given a set of data points (outcomes). The Log-Likelihood function is important in statistical inference, and it is commonly used to estimate a parameter from a set of statistics. Log likelihood is usually a negative value, and a larger LL value indicates a better model.

4. Finally, the observed value is plotted against the predicted value. And the relative error is also provided for model selection.

The overall model performance comparison is given in Table 24, the results suggest Poisson model dominant negative binomial model by producing smaller AIC and deviance and larger Log-likehood. Figure 13 and 14 plot the observed pickups against the predicted value. Ideally, if all the scattered dots fall into the 45 degree straight line indicate perfect prediction power. The prediction of both models produce some noise due to high variance in the training dataset in these two figures. Furthermore, Figure 15 and 16 plot the relative error calculated as: Relative error = (Prediction value - Observed value)/Observed value. From the figures, we observe that the relative errors are evenly distributed around the red line. Overall, both negative binomial regression and Poisson regression models produces fairly good prediction. However, since Poisson distribution is a discrete probability distribution that expresses the probability of a given number of events occurring in a fixed interval of time or space if these events occur with a known constant rate and

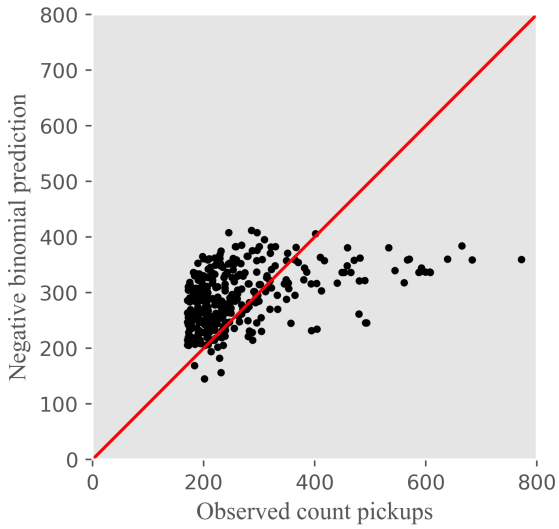

Figure 13. Negative binomial regression prediction vs. observed pickups

independently of the time since the last event [82], Poisson regression model is selected due to better interpretation of a random taxi pickup.

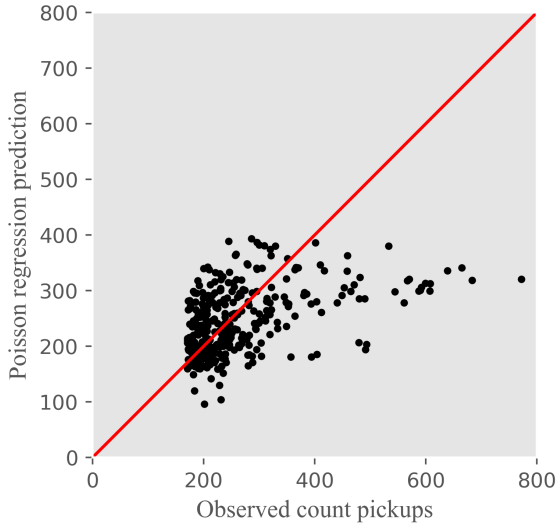

Figure 14. Poisson binomial regression prediction vs. observed pickups

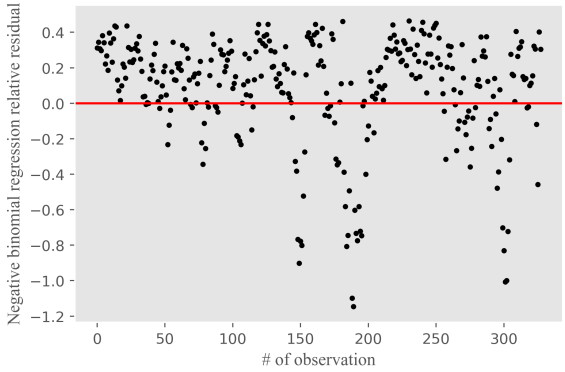

Figure 15. Negative binomial regression relative error

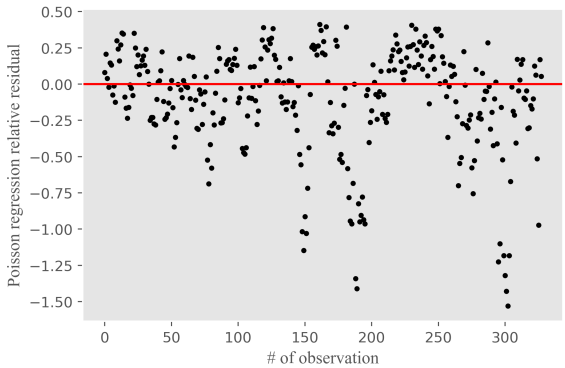

Figure 16. Poisson regression relative error

5.2.3 Supply forecasting

In this section, we aim to understand when and where taxis become available for next customer after a dropoff. Without loss of generality, we make the following assumptions.

1. First, the taxi supply in this dissertation is defined as the number of drop-offs in a neighborhood during a period of time, specifically in hourly interval. This is done because number of drop-offs, in a relatively short period of time, is an indicator of how many empty taxis become available in that census tract.

2. Second, as soon as the drop off occurred, the taxi immediately becomes available for next customer.

3. The taxi drivers are assumed to be fully aware of the hot spot of picking up customers thus they will search in such nearby popular neighborhood or even drive closer to the hot spots if previous drop-off location is away from these spots.

We perform time series analysis to first observe the variation of taxi dropoff within hourly interval and study the drop offs trend over time as shown in Figure 17. Figure 17 plots the weekly counts of drop offs for census tract Manhattan 5200 from April 21 to 26. Two main findings can be summarized from the time series plot. First, the weekly count of drop offs display strong seasonality behavior. Often, seasonality is defined to be the tendency of time-series data to exhibit behavior that repeats itself every L periods [83]. The term season is used to represent the period of time before behavior begins to repeat itself. In Figure 17, for example, the weekly time series plot clearly shows that the count of drop offs reach peak between 8:00 am to 9:00 am, and then hit the bottom around midnight almost for every weekday. This pattern may repeat from week to week, therefore, the season length of period can be concluded as 24 hours. Second, Figure 17 also shows that the total drop offs

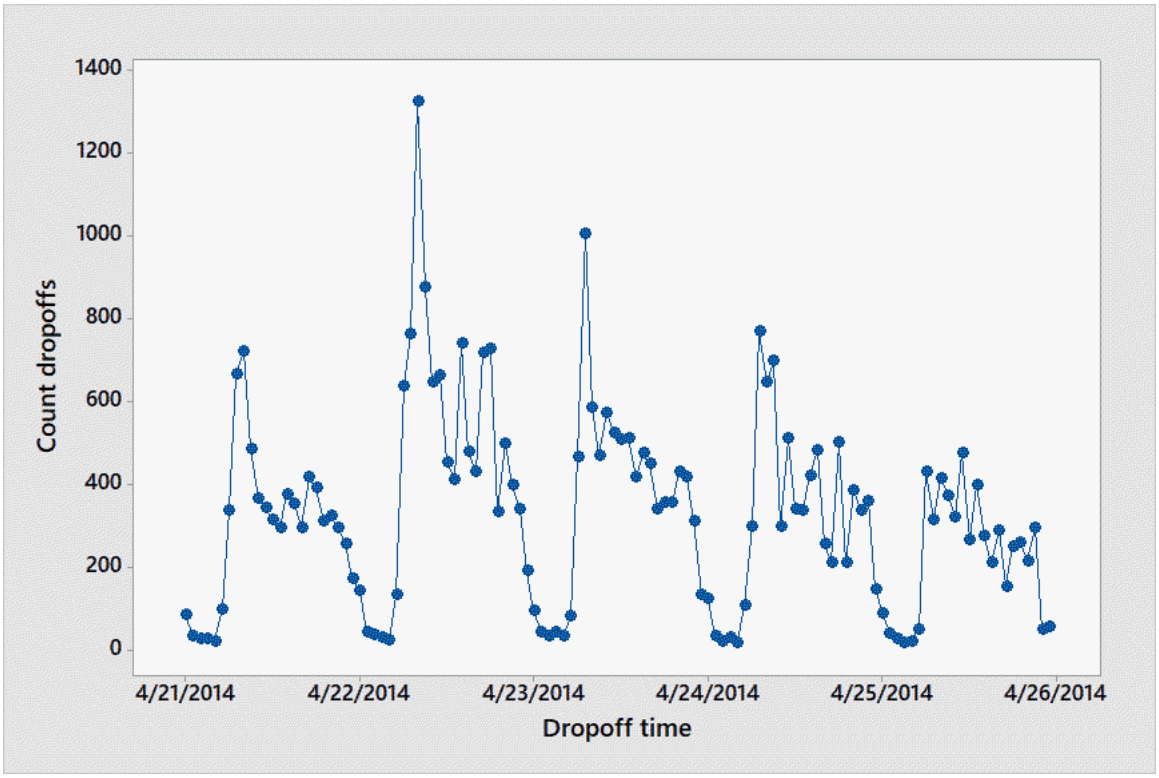

Figure 17. Hourly dropoffs in census tract Manhattan 5200

drifting upward from Monday to Tuesday and then the overall trend decrease from Wednesday to Friday. Thus stationary process is required before performing time series forecasting.

There are two main stationary process can be found in the literature [84]:

Definition 5. *First Order Stationary: A time series is a first order stationary if expected value of X(t) remains same for all t.*

For example in economic time series, a process is first order stationary when one remove any kinds of trend by some mechanisms such as differencing.

Definition 6. *Second Order Stationary: A time series is a second order stationary if it is first order stationary and covariance between X(t) and X(s) is function of length (t-s) only.*

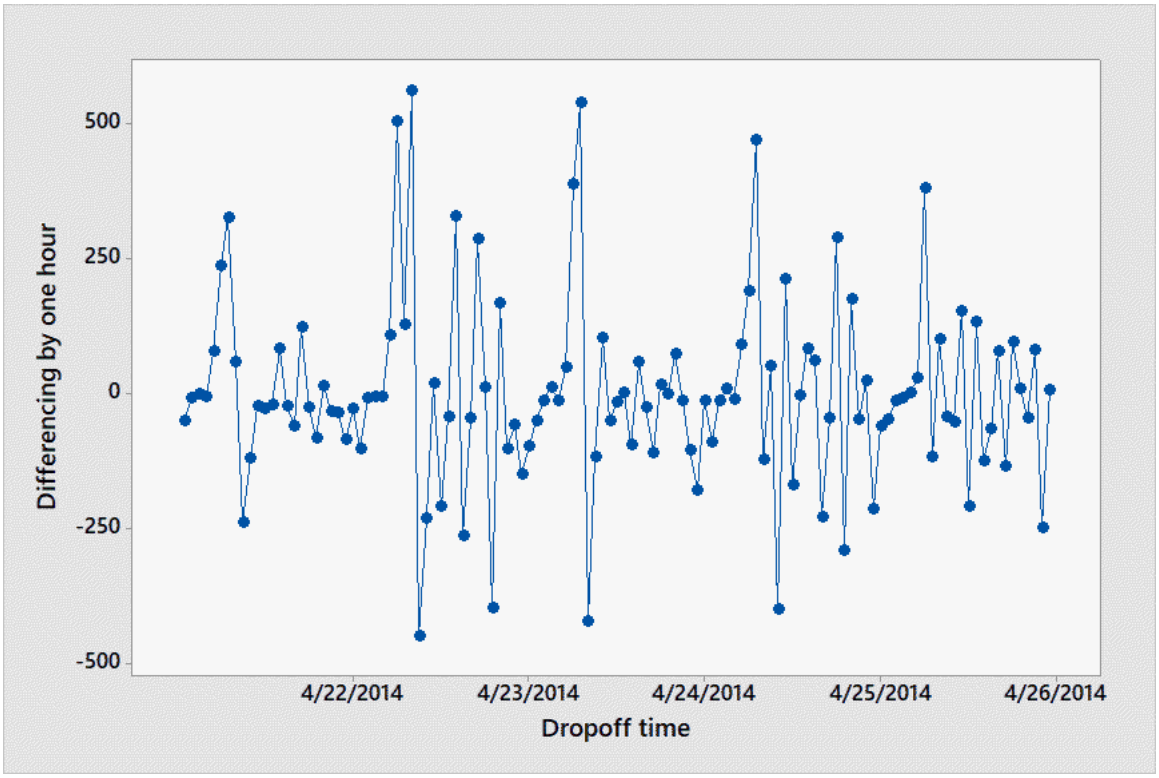

Figure 18. Differencing by 1 hour dropoffs in census tract Manhattan 5200

A second stationary process can be found when one stabilizes the time series by transformation of time dependent variables, such as squared value or logarithms.

In time series forecasting, the differencing term can be determined by computing the changes between the successive observations for a given time period until the time series plots become stationary. In most cases, the first order difference is usually adequate enough to obtain stationary as shown in Figure 18. Once the non-stationary has been fixed , we are then able to select appropriate time series regression model to develop the supply forecasting model. One of the most commonly used time series forecasting is known as Box-Jenkins method, as known as ARIMA. There are three primary parameters to determine in the ARIMA models: the degree of auto regressive term, differencing order and degree of moving average term.

To determine a proper ARIMA model, we first check the partial correlation

and auto correlation. The auto correlation coefficient (ACF) is given as:

$$r_k = \frac{\sum_{i=1}^{N-k}(Y_i - \bar{Y})(Y_{i+k} - \bar{Y})}{\sum_{i=1}^{N}(Y_i - \bar{Y})^2}, \tag{43}$$

where Y_i denote to the measurements at time X_i and k refers to lag degree. While the partial correlation coefficient (PACF) is given by:

$$\frac{\text{Covariance}(y_i, x_{i-h}|x_{i-1}, ..., x_{i-h+1})}{\sqrt{\text{Variance}(y_i|x_{i-1}, ..., x_{i-h+1})\text{Variance}(y_{i-h}|x_{i-1}, ..., x_{i-h+1})}}, \tag{44}$$

For a particular time series, the h^{th} order partial auto correlation is the partial correlation of y_i with y_{i-h}, conditonal on y_{i-1} ,..., y_{i-h+1} as in equation 44. Figure 19 and 20 plot the ACF and PACF for various lag orders. First, both the ACF and PACF present small correlations across all the lags, this indicates that the time series may not need higher order of differencing thus the degree of ordering can be set as 1. Second, negative correlation on ACF and PACF suggest both AR and MA terms should be included in the model. Third, the order for those terms can be determine by the lag beyond which the ACF/PACF cuts off. In our case, both differecing degree and order of AR/MA terms are set to 1 as those preliminary analysis suggests. Once the degree of differencing and the order of AR/MA terms has been calculated and the behavior of them has been examined, the next step is to use the time series data to estimate the tentative model. In theory, the parameters of the selected model can be generated through least squares. However, nonlinear least squares algorithms usually consist of a combination of search routines, i.e., various combination of different order of AR/MA terms. Then it needs to be implemented through an iterative process before finalizing the model [84]. In order to select the optimal combination of AR and MV terms, we evaluated the hyperparameters of ARIMA model by grid search method with various combination of different degrees of auto regressive term, moving average term and differencing order. Based on previous differencing residual plot and ACF/PACF analysis, we set the range for each parameter as [1,2,3], which gives us 27

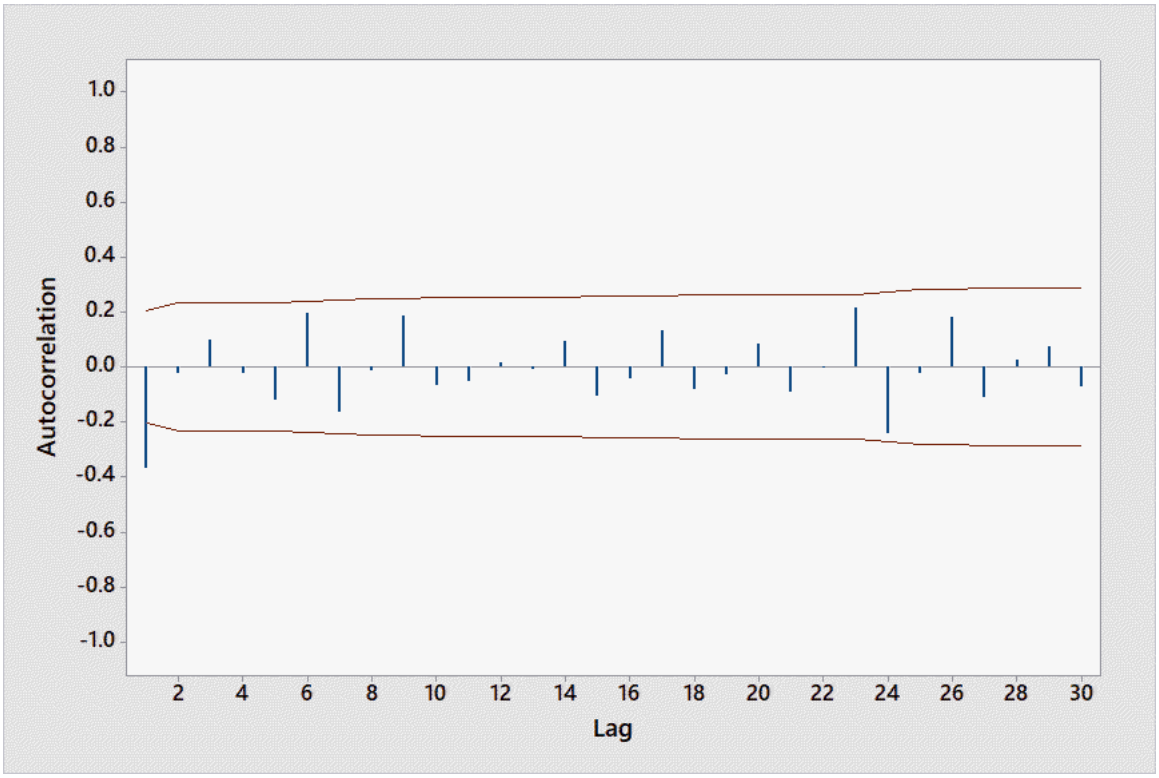

Figure 19. ACF for residuals for Drop offs

combinations in total. The main steps are:

1. The data is split to testing and training set and all the ARIMA models are developed based on the training set.

2. After one ARIMA model is trained, the prediction is then stored and compared with actual observations.

3. The mean squared error is calculated as: $\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2$.

Finally, the overall MSE across all possible combination is given in Table 25. The lowest MSE was obtained by ARIMA model (1,1,1) at 10490.7.

The ARIMA model parameters are reported in Table 26, the zero p-values for each term shows those parameters are statistically significant from zero. However, as any other regression techniques, the diagnostics step need to be performed before

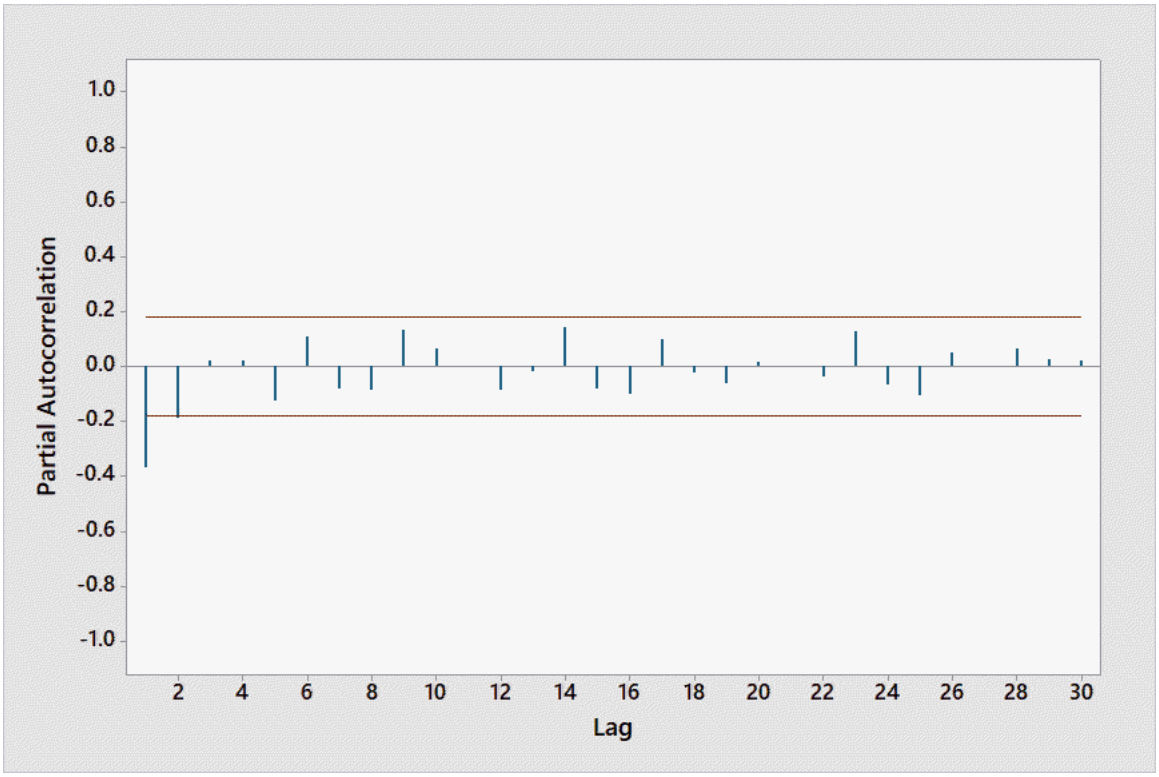

Figure 20. PACF for residuals for Drop offs

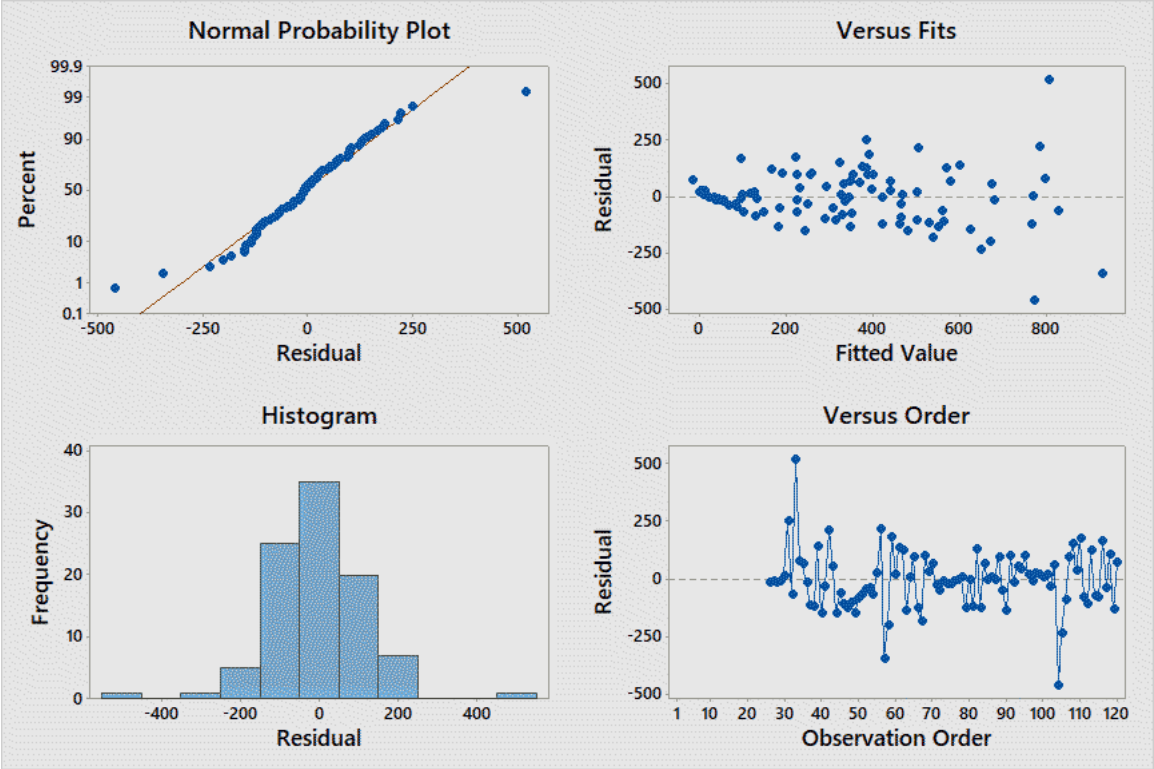

Figure 21. Diagnosis for ARIMA(1,1,1)

the model implementation. The diagnostic analysis is carried out to validate the model, or to check if the tentative model needs modification. Figure 21 plots the normal probability and histogram of residuals, as well as residuals against the fitted value and observation order, respectively. First, the normal probability plot and histogram confirm that the residuals from ARIMA (1,1,1) model are approximately randomly distributed and the residuals against fitted values and observation order show that the residuals are independent of each other.

In additional to ARIMA, another widely used time series regression technique, Holt-Winter's [85] method was also tested by the sample weekly dropoffs data. To deal with seasonality in time series, two different type models were proposed by Winters: additive seasonality and multiplicative seasonality. The former refers to the seasonal trend (increase or decrease) are independent to previous time periods. For example, the demand of a certain product for each

month may increase by 100. Thus, we could add to our forecasts for every month by the amount of 100 over the respective monthly average to account for this seasonal fluctuation. In this case, the additive model is appropriate for this type of time series prediction. Alternatively, during the same time period, if the demand for the same product may increase or decrease by 20%. In this case, the increment is measured by a factor of 1.2. Thus, when the demand is small, then the absolute increase for each month will be relatively weak, but with the same percentage. Whereas if the demand are sufficiently large, then the absolute increase in dropoffs will be proportionately greater. In such case, the nature of the seasonality is multiplicative and thus the regression model should be selected accordingly.

Recall from Figure 17, instead of an absolute increment on dropoffs, the time series displays a strong multiplicative seasonality by a certain factor from Monday to Friday, thus the multiplicative Holt-Winters method is selected. Below is the general equations for multiplicative Holt-Winter's model [85],

$$\hat{Y}_{t+h|t} = (\ell_t + hb_t)s_{t+h-m} \tag{45}$$

$$\ell_t = \alpha \frac{y_t}{s_{t-m}} + (1-\alpha)(\ell_{t-1} + b_{t-1}) \tag{46}$$

$$b_t = \beta(\ell_t - \ell_{t-1}) + (1-\beta)b_{t-1} \tag{47}$$

$$s_t = \gamma \frac{y_t}{(\ell_{t-1} + b_{t-1})} + (1-\gamma)s_{t-m}, \tag{48}$$

where $\hat{Y}_{t+h|t}$ is the forecast for time period $t+h$, ℓ_t is the estimation of deseasonalized level, b_t is the estimate of the overall trend , s_t is the seasonal component estimation and m is the seasonal period.

In Equation 46, ℓ_t is the overall smoothing and α is the smoothing constant between 0 and 1. The seasonal factor for time period T is calculate as dividing y_t by s_{t-m}. This step is to deseasonlize the time series data such that there is only the trend component and the prior value to update ℓ_t. The smoothing of the trend factor is giving in Equation 47. It is simply the smoothed difference between two

successive estimates of the deseasonalized level. β is the second smoothing constant between 0 and 1.

Finally, the seasonal index is given by s_t. It is a combination of the most recently observed seasonal, factor given by the demand y_t divided by the deseasonalized series level estimate ℓ_t and the previous best seasonal factor estimate for this time period. γ is the third smoothing constant and between 0 and 1.

To determine proper values for the three smoothing constant α, β and γ, a nonlinear programming is formulated to minimize the mean square error as:

$$Minimize \qquad \frac{1}{n}\sum_{i=1}^{n}(\hat{Y}_{t+h|t} - Y_{t+h|t})^2 \tag{49}$$

$$Subject\ to \qquad \alpha, \beta, \gamma \in (0, 1), \tag{50}$$

where $\hat{Y}_{t+h|t}$ is estimated from Equation 45 with initial α, β and γ all set to 0.1 and time periods t, h set to 96 hours and 24 hours, respectively. The optimal value for each parameter is reported in Table 27.

As for the ARIMA model, the diagnostic analysis for Holt-Winter's method is given in Figure 22. Normality of residuals are confirmed by plotting the normal probability and histogram, the residuals are approximately normality distributed and independent of each other by plotting the residuals against fitted value and observation order.

Finally, we use three measures to compare the ARIMA and Holt-Winter's models. They include: the mean absolute percentage error (MAPE), median absolute deviation (MAD) and mean squared deviation (MSD). The comparison are given in Table 28. It clearly shows that Holt-Winter's method outperforms ARIMA with 19.95% of MAPE, 70.24 of MAD and 9978.31 of MSD Thus, Holt-Winter's multiplicative model is selected.

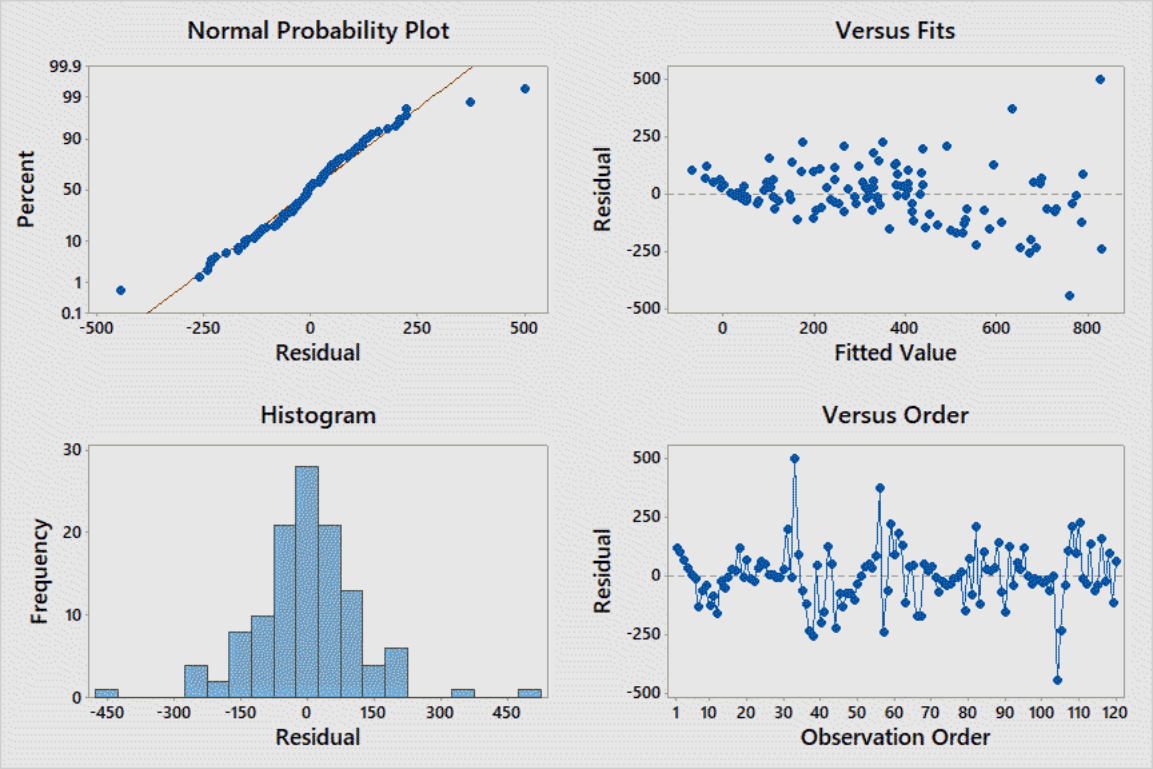

Figure 22. Diagnosis for Winter's method

5.3 Computational Results for the Integrated two-stage Uber-like Ridesharing Assignment Models

In this section, we integrate the two-stage uber-like ridesharing assignment models with the supply and demand forecasting models and test the integrated models using NYC real-life data introduced in the previous sections. In particular, we use these real data to evaluate the proposed multi-criterion optimization model (resource allocation problem) and the individual driver decision support model (prospect maximization problem). Both the multi-criteria optimization model and individual user model are implemented and solved in GAMS [80], a state-of-the-art modeling language for nonlinear programs. Particularly, the first-stage welfare maximization problem is solved by CPLEX and the second-stage users' prospect maximization is solved by MNOS. All simulations were run on a 16-core dual Opteron CPU server with 32GB of memory running openSUSE 11 Linux.

The data used in our numerical simulation is organized in three sets. First, the population size of passengers and drivers are generated from the predictive models developed in

Section 5.2.2 and section 5.2.3, after a certain census tract of interest is selected. Second, without loss of generality, we assume the spatial locations of participants are distributed in three different traffic zones that are corresponded to different traffic volume. The distance between each passenger and driver are replaced by $D_{i,j} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}$. The gender of passengers and drivers and the review rating for drivers are uniformly distributed. Lastly, the monetary value per mile is given in Table 20. These considerations lead to the following outline of a random instance generator.

Step 1: Determine the census tract of interest, then apply the forecasting model by gathering the census tract demographic information and hourly dropoffs, to generate the number of passenger and drive, i.e., I, J as in Section 3.4.4.

Step 2: Create three traffic zones representing high, medium and low traffic volume representing different traffic condition (with high traffic volume indicating increased travel time and operational cost). The corresponding radius of the three circular traffic zones are $r1$, $r2$, $r3$, respectively.We distribute 40% of total requests within the high traffic volume zone, 40% within the medium and 20% within the low traffic volume zones, respectively.

Step 3: Generate the drivers' current location (i.e., longitude and latitude coordinates) information. The drivers distribution follows the same distribution when generate three traffic zones in step 2, i.e., 40% in high traffic volume zone, 40% in medium traffic volume zone, 20% in low traffic volume zone.

Step 4: Generate the drivers' current location (i.e., longitude and latitude coordinates) information. For each request origin, the distance between origin and destination are generated from a uniform distribution $y \sim U$ [2,20].. The destinations are scattered distributed and the angle between each origin and destination is calculated as: angle(i) = max(0, angle(i-1)+$2\pi/I$), where I is the number of requests.

We conduct the model validation using the examples displayed in Table 29. It consists of five census tracts including 5400, 5200, 100, 1900 and 1500 from three different Boroughs: Manhattan, Queens and Brooklyn. The drivers supply and passengers demand were forecasted using the models developed in section 5.2.2 and 5.2.3. The heat map representing each census tract with scaled density of demand is shown in Figure 23.

Figure 24 provides a closer look on how passenger and driver locations distributed. The trip information for the 360 passengers and 110 drivers are generated for census tract 5400 at 5:00 pm. First, the coordinate information for each passenger and driver is generated within the census tract geographic boundary. Second, the gender information for each passenger and driver is randomly generated

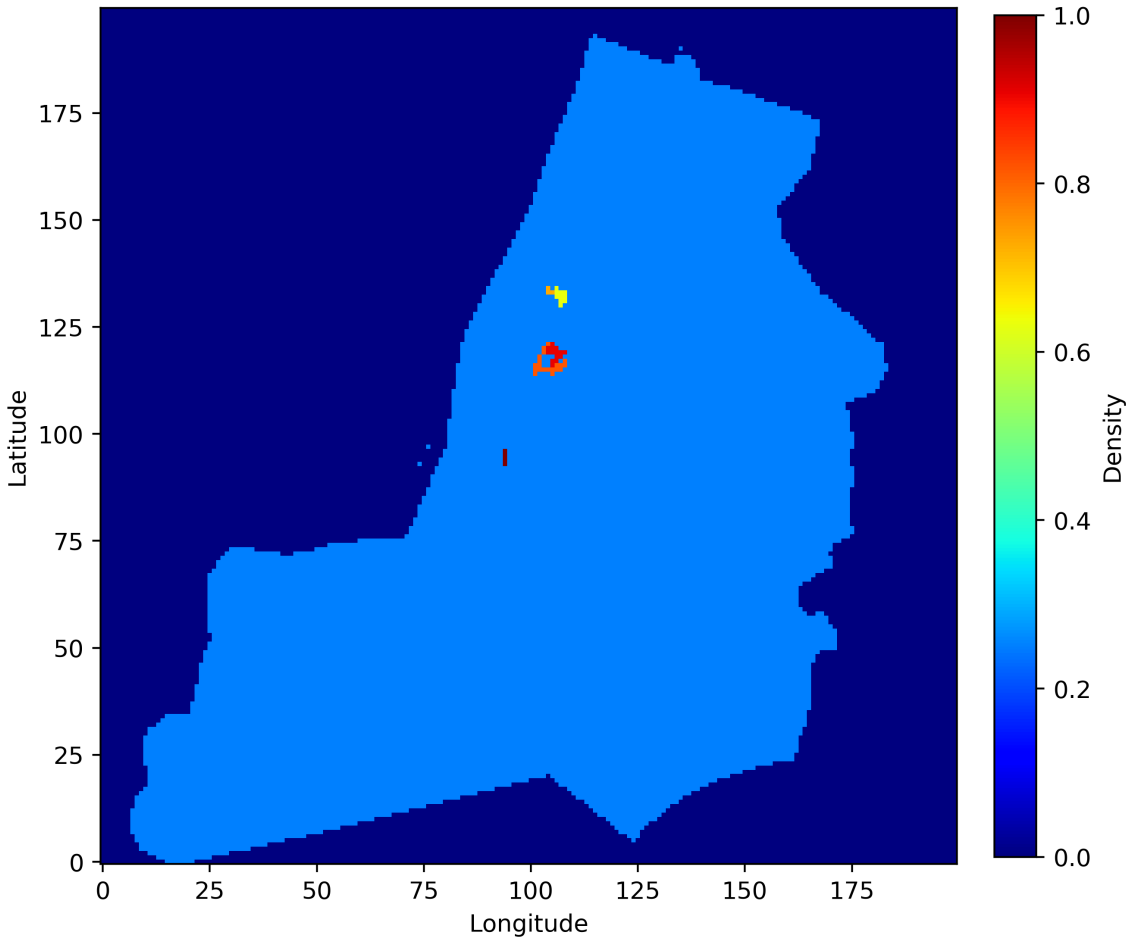

Figure 23. Five census tracts with scaled density plot

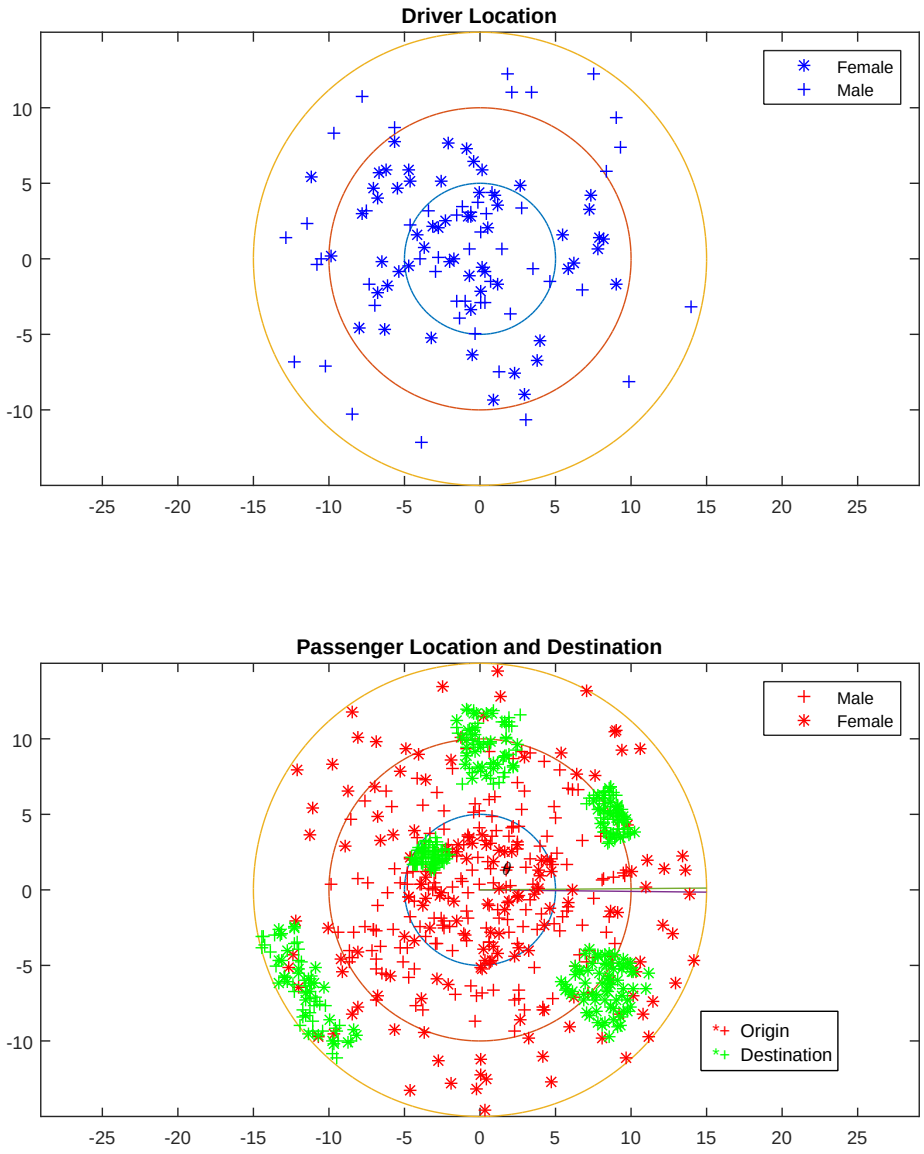

Figure 24. Passenger and driver distribution in census tract Manhattan 5400

from a uniform distribution $x \sim U$ [0,1] and the driver's review rating is also randomly generated from a uniform distribution $x \sim U$ [0,5]. The profit for each potential driver-passenger pair (passenger i and driver j) is calculated as: $P_{i,j} = D_i R - (L_{i,j} + D_i)G$, where D_i is the total travel distance from passenger i's origin to his/her destination, $L_{i,j}$ is the distance between driver j's current location and passenger i's current location. R, G are the fare rate generated from Section 5.2.1 and operation costs (dollar per mile), respectively. In general, we solve the problem in two stages. As mentioned previously, in the first stage we solve a multi-criteria problem that considers three objectives including gender matching maximization, review rating maximization and profit maximization. We solve three problem sequentially using lexicographic algorithm. In the second stage, the optimal solution from Stage 1 is used by Stage 2 as an initial assignment, where each driver is bidding on their own behalf and their bidding strategy is restricted by certain lower and upper bounds. That is a driver is not willing to lose money in order to win a particular task and the bidding price cannot exceed a upper bound to avoid overcharge. After a decentralized driver prospect maximization is solved in Stage 2, the winners are announced immediate and the rests of those passengers/drivers not being served/assigned will go to next iteration until either passenger or driver is fulfilled. The solution algorithm is summarized as follows:

Stage 1: *Tentative driver-passenger assignment*

Input: I - passenger set, J - driver set:
Output: P - tentative driver-passenger assignment
for $c \in C$ **do**:
    Solve max F_c $|s.t$ (4)-(10) $\rightarrow F_c = F_c^*$
    add constraint $F_c = F_c^*$
end for

Stage 2: *Driver auction*

Input: P - tentative driver-passenger assignment
Output:F - final driver-passenger assignment
for $j \in P$ **do**:
    Solve max Z_j $|B_{i,j} \in]LB, UB[$
end for
if $I \leftarrow \emptyset / J \leftarrow \emptyset$ **then**
    Stop

Elseif $I \neq \emptyset$ & $J \neq \emptyset$ **then**
    Remove F from I,J and go to Stage 1.
end if

Table 30 to Table 33 summarize the overall performance for the simulated cases for census tract Manhattan 5400, Manhattan 5200, Queens 100, Queens 1900 and Brooklyn 1500. First,our proposed model ensures that all the requests are fulfilled after the iterative process. In each table, column "Fulfilled requests " records the satisfied requests while columns "Gender matching ", "Average rating" and "Average profit" report the three objective values by solving the multi-criteria optimization model in Stage 1. Column "Average prospect" reports the average prospect after solving the prospect optimization problem for the winning drivers in the bidding process during Stage 2. Under each column, two sub columns "Per

iteration" and "Cumulative" that report the corresponding performance measure for each iteration. Several observations can be made from these two tables.

First, we note that during the iteration process, the average gender matching rate decreases for all five census tracts. This is due to the decreased availability of female drivers as the iteration proceeds further. The gender matching rate eventually decreases to zero indicating no female drivers are available. Cumulatively, the lowest average gender matching rate among all five census tracts is 71%. This indicates at least more than %70 of the passengers can be severed with the same gender driver.

Second, besides the gender matching rate decreasing, the average review rating of assigned driver also decreases for each census tract as shown in Figure 26. Since the objective function for female-passenger-to-female-driver matching rate has the highest priority, the system has to compromise by selecting female drivers with lower review rating to satisfy female-passenger-to-female-driver matching rate from the first objective function. Furthermore, from Figure 27, we also observe that the cumulative average profit also decreases but the difference between highest and lowest is very marginal.

Lastly, the performance of the proposed model in stage 2 model yields average prospect 3, 3.05, 2.94, 2.87 and 2.80 for five cases respectively. Figure 28 plots the cumulative average prospect for all the assigned drivers across all the simulated cases during the iteration process. This clearly shows that the prospect is oscillating from 2 to 3. This is because the prospect is only determined by the driver's bidding strategy and it is irrelevant to the size of the problem. Therefore the appropriate price scheme is very important to help the drivers to identify and evaluate the most desirable passenger request.

Table 30 explicitly reports the overall performance for each step for census tract Manhattan 5400. All the 110 requests are fulfilled after 29 iteration steps. In the first 5 iterations, the average gender matching strictly equals to 1 because of

sufficient female drivers. The gender matching rate then decreases after iteration 5 as the number of available female drivers decreases. Similar pattern can be observed for the average driver's review rating and average profit per pair, the review rating starts with 4.49 and finally drops to 3.52 while the average profit decreases from 58.78 to 54.25. However, due to different priority in the solution algorithm, the decreasing slope for each objective is ranking in descending order as: gender, review and profit. Finally, all the winning driver's prospect is achieved at 3.00 on average.

Similarly, Table 31 summarizes the simulation results in census tract Manhattan 5200. The generated problem size is slightly smaller than Manhattan 5400, due to lower average household income and smaller population. Therefore, it takes 26 iteration steps to complete the passenger's request. The model behavior is similar to the case for Manhattan 5400. Gender matching rate stays at 1 for the first five iterations before it declines due to reduced availability of female drivers. Also similar to the previous case, the average maximal prospect for all drivers ranges between 2 and 3.05.

Among all the simulated cases,the census tract Queens 100 has the highest forecasted supplies (390) and demands (143) due to its largest population size. Our computation takes 36 iterations to have all passenger requests fulfilled. Table 32 displays the output for all the three objectives in Stage 1 and the average prospect in Stage 2, respectively. Since the problem size increases, the number of available female drivers increases, thus we expect more female passenger's requests can be fulfilled by female drivers, as the gender matching rate constantly equals 1 until iteration 7. And the average review rating and profit starts with 5, 39.65 and then ends up with 3.81 and 36.96, respectively.

Lastly, Table 33 and 34 show the results for the two simulated cases with smallest problem size among the five census tracts: Queens 1900 and Brooklyn 1500. The computational complexity is slightly reduced. Both of those two cases spend less than 25 steps to complete the multi-criteria optimization problem and

driver's prospect maximization problem. Gender matching rate decreases from 1 to 0.7099, average review rating starts with 5 and ends at 3.62, average profit for early assigned pairs can reach up to 41 and finally reach to 38.59 on average for those pairs towards the end of the process.

Finally, Figure 27 presents a interesting phenomenon that the average profit in Manhattan is obviously higher than the other three census tracts. This can be interpreted as follows. First, the passenger's origin and destination pairs tend to have longer travel distance on average in these two census tracts. Second, the driver to passenger ratios is relatively lower than other census tract, hence the driver's bidding strategy is more aggressive comparing to other cases.

Overall, the performance of our proposed model from the simulated experiments for those selected census tracts can be summarized as follows:

1. The proposed model generates satisfactory assignment to the passenger's request: the lowest gender matching rate at 70% can resolve most of the female passenger's safety concerns when choosing crowdsourcing transportation platform. And the lowest average drivers review rating at 3.65 ensures that only drivers with above average review rating to be selected.

2. The proposed model optimize two attributes from the driver's perspective: the profit for each assignment from multi criteria maximization model in Stage 1 and the prospect maximization in Stage 2. The optimized profit ensure each assignment is profitable and the maximize prospect provides a decision support means for drivers when under multi tasks selection scenario.

TABLE 19

Data preparation for raw taxi data

	Criteria	**Removed records**
Step1	Remove pickups outside NYC	322,789
Step2	Remove duplicate/missing value	75,035
Step3	Remove false trip records	137,195
Overall	**3% of raw data was removed**	535,020

TABLE 20

Linear regression results

	Sample size	R^2	MSE	Regression model
Rush ours dataset	8,443,156	0.9044	3.53	y = 5.45 + 3.28x
Non rush hours dataset	5,640,583	0.9055	3.85	y = 5.11 + 3.34x

TABLE 21

Sample input data for census tract Manhattan 5400

Population = 4,536
Number of employed people = 2,603
Mean household income = $227,485

Hour	Monthly Pickups	Hour	Monthly Pickups
0	9,124	12	20,165
1	5,214	13	19,945
2	3,398	14	21,230
3	2,302	15	21,016
4	2,698	16	18,789
5	4,289	17	23,355
6	10,903	18	27,979
7	18,658	19	27,351
8	22,225	20	25,883
9	21,909	21	23,936
10	19,298	22	21,126
11	19,062	23	16,013
Total		405,868	

TABLE 22

Poisson regression results

	coef	**std err**	**z**	**P>|z|**	**[0.025**	**0.975]**
Intercept	5.3222	0.009	618.196	0.000	5.305	5.339
Population	7.045e-05	6.29e-06	-11.197	0.000	-8.28e-05	-5.81e-05
Avg_income	9.331e-07	4.95e-08	-18.860	0.000	-1.03e-06	-8.36e-07
NumofEmployed	1.841e-05	9.34e-06	1.971	0.049	1.07e-07	3.67e-05
5PM	0.5926	0.009	62.811	0.000	0.574	0.611
6PM	0.7272	0.008	92.114	0.000	0.712	0.743
7PM	0.7104	0.008	87.457	0.000	0.694	0.726
8PM	0.6651	0.008	80.835	0.000	0.649	0.681
9PM	0.6178	0.009	70.209	0.000	0.601	0.635
10PM	0.6001	0.010	60.984	0.000	0.581	0.619
8AM	0.7451	0.011	65.434	0.000	0.723	0.767
9AM	0.6641	0.011	58.400	0.000	0.642	0.686

TABLE 23

Negative binomial regression results

	coef	**std err**	**z**	**P>|z|**	**[0.025**	**0.975]**
Intercept	5.3134	0.049	107.669	0.000	5.217	5.410
Population	6.698e-05	3.22e-05	-2.079	0.038	-0.000	-3.84e-06
Avg_income	9.647e-07	2.64e-07	-3.652	0.000	-1.48e-06	-4.47e-07
NumofEmployed	1.882e-05	4.8e-05	0.392	0.695	-7.52e-05	0.000
5PM	0.6040	0.050	12.041	0.000	0.506	0.702
6PM	0.7268	0.044	16.559	0.000	0.641	0.813
7PM	0.7038	0.045	15.546	0.000	0.615	0.793
8PM	0.6633	0.045	14.752	0.000	0.575	0.751
9PM	0.6174	0.047	13.035	0.000	0.525	0.710
10PM	0.5997	0.052	11.455	0.000	0.497	0.702
8AM	0.7363	0.061	12.098	0.000	0.617	0.856
9AM	0.6621	0.060	11.019	0.000	0.544	0.780

TABLE 24

Model performance comparison

Model	AIC	Deviance	Log-likehood
Poisson regression	37,465	28.7	-2156.1
Negative binomial regression	38,996	897.6	-5693.4

TABLE 25

ARIMA optimal parameter

(p,d,q)	SE	(p,d,q)	SE
(1, 2, 3)	11722.3	(2, 1, 3)	20132.3
(1, 2, 1)	11510.6	(2, 1, 1)	13639.1
(1, 2, 2)	11126.6	(2, 2, 3)	11187.0
(1, 3, 1)	10544.8	(2, 2, 1)	13882.1
(1, 3, 2)	12706.2	(2, 2, 2)	13490.1
(1, 3, 3)	24627.3	(3, 1, 2)	11593.7
(1, 1, 2)	13101.5	(3, 1, 3)	12949.0
(1, 1, 3)	11735.2	(3, 1, 1)	19625.2
(1,1,1)	***10490.7***	(3, 2, 3)	13752.6
(2, 3, 1)	13895.7	(3, 2, 1)	22303.7
(2, 3, 2)	13944.3	(3, 2, 2)	22316.4
(2, 3, 3)	13345.0	(3, 3, 1)	22338.5
(2, 1, 2)	12209.1	(3, 3, 2)	22322.4
(2, 1, 3)	20132.3	(3, 3, 3)	22371.8

TABLE 26

ARIMA model results

Estimates of Parameters ARIMA (1,1,1)

Type	Coef	SE Coef	T-Value	P-Value
SAR 24	-0.415	0.110	-3.78	0.000
SMA 24	0.8515	0.0965	8.82	0.000
Constant	-47.97	3.19	-15.06	0.000

TABLE 27

Holt-Winter's optimal parameter

Holt-Winter's optimal parameter	
α	0.3395
β	0.3700
γ	0.1564

TABLE 28

Model performance comparison

Accuracy measures	MAPE	MAD	MSD
Holt-Winter's method	19.95%	70.24	9978.31
ARIMA	25.95%	74.67	10490.69

TABLE 29

Five census tracts with forecasting demand and supply

Census tract	Borough	People	Avg household income	# of employee	Forecasting drivers	Forecasting passengers
5400	Manhattan	4536	$227485	2603	360	110
5200	Manhattan	3726	$132813	3101	319	97
100	Queens	6430	$120807	4250	390	143
1900	Queens	1655	$117788	1187	263	76
1500	Brooklyn	2562	$115000	1013	273	79

TABLE 30

Simulated performance for census tract Manhattan 5400

	Gender matching		Review rating		Avg profit		Avg prospect		Fulfilled requests	
Iteration	Per Iteration	Cumulative	Per Iteration	Cumulative	Per Iteration	Cumulative	Per Iteration	Cumulative	Per Iteration	Cumulative
1	1	1	4.49	4.49	58.75	58.75	3	3.00	10	10
2	1	1	4.36	4.43	58.51	58.64	3.46	3.22	9	19
3	1	1	4.2	4.36	58.28	58.53	3.25	3.23	8	27
4	1	1	4.08	4.30	57.77	58.37	3.11	3.20	7	34
5	1	1	4.07	4.26	57.5	58.22	2.56	3.09	7	41
6	0.9247	0.9904	4	4.23	57.3	58.11	3.32	3.12	6	47
7	0.8894	0.9790	3.99	4.20	56.85	57.96	3.25	3.14	6	53
8	0.8462	0.9675	3.81	4.17	56.31	57.82	2.79	3.11	5	58
9	0.84	0.9593	3.72	4.14	55.61	57.68	2.7	3.08	4	62
10	0.7943	0.9493	3.54	4.10	54.89	57.51	2.65	3.05	4	66
11	0.7924	0.9403	3.48	4.07	53.84	57.30	3.1	3.06	4	70
12	0.7431	0.9297	3.39	4.03	52.79	57.06	3.16	3.06	4	74
13	0.7253	0.9217	3.36	4.00	52.63	56.88	2.92	3.06	3	77
14	0.6987	0.9133	3.23	3.98	51.38	56.68	2.62	3.04	3	80
15	0.688	0.9052	2.98	3.94	51.35	56.49	2.82	3.03	3	83
16	0.6811	0.8974	2.97	3.91	50.35	56.27	3.15	3.04	3	86
17	0.6147	0.8909	2.94	3.88	50.32	56.14	3.03	3.04	2	88
18	0.5966	0.8844	2.92	3.86	49.75	55.99	2.59	3.03	2	90
19	0.52	0.8765	2.67	3.84	49.37	55.85	3.27	3.03	2	92
20	0.5067	0.8686	2.47	3.81	49.23	55.71	2.54	3.02	2	94
21	0.4741	0.8604	2.45	3.78	49.07	55.57	2.5	3.01	2	96
22	0.4491	0.8520	2.24	3.75	48.53	55.43	2.96	3.01	2	98
23	0.4335	0.8436	2.14	3.72	47.74	55.27	2.68	3.00	2	100
24	0.3646	0.8342	2.08	3.68	47.13	55.11	3.39	3.01	2	102
25	0.3106	0.8242	1.49	3.64	44.69	54.91	2.96	3.01	2	104
26	0.2755	0.8138	1.49	3.60	43.24	54.69	2.8	3.01	2	106
27	0.2088	0.8026	1.45	3.56	42.44	54.47	2.62	3.00	2	108
28	0	0.7953	1.31	3.54	42.21	54.35	3.19	3.00	1	109
29	0	0.7880	1.08	3.52	42.01	54.24	2.56	3.00	1	110

TABLE 31

Simulated performance for census tract Manhattan 5200

	Gender matching		Review rating		Avg profit		Avg prospect		Fulfilled requests	
Iteration	Per Iteration	Cumulative	Per Iteration	Cumulative	Per Iteration	Cumulative	Per Iteration	Cumulative	Per Iteration	Cumulative
1	1	1	4.71	4.71	44.79	44.79	3.08	3.08	9	9
2	1	1	4.69	4.70	44.69	44.74	2.84	2.97	8	17
3	1	1	4.59	4.67	44.65	44.71	2.89	2.94	8	25
4	1	1	4.57	4.64	44.55	44.68	3.16	2.99	7	32
5	1	1	4.32	4.59	43.75	44.51	3	2.99	7	39
6	0.9495	0.9943	4.28	4.55	43.57	44.40	3.14	3.01	5	44
7	0.9058	0.9852	3.84	4.48	43.13	44.27	3.15	3.02	5	49
8	0.9008	0.9774	3.5	4.39	43.12	44.17	3.17	3.04	5	54
9	0.8621	0.9695	3.38	4.32	42.84	44.08	3.06	3.04	4	58
10	0.7767	0.9570	3	4.23	42.07	43.95	3.05	3.04	4	62
11	0.7423	0.9440	2.97	4.16	41.95	43.83	3.29	3.05	4	66
12	0.6867	0.9328	2.87	4.10	41.75	43.74	3.11	3.06	3	69
13	0.6563	0.9213	2.81	4.05	41.56	43.64	2.88	3.05	3	72
14	0.5984	0.9084	2.8	4.00	41.19	43.55	3.18	3.05	3	75
15	0.5956	0.8964	2.75	3.95	41.04	43.45	3.02	3.05	3	78
16	0.5622	0.8880	2.64	3.92	41.03	43.39	3.01	3.05	2	80
17	0.5017	0.8786	2.35	3.88	40.11	43.31	3.01	3.05	2	82
18	0.4665	0.8688	2.22	3.84	40.05	43.23	3.12	3.05	2	84
19	0.4399	0.8588	1.86	3.79	39.63	43.15	3.18	3.06	2	86
20	0.4375	0.8492	1.79	3.75	39.49	43.06	3.21	3.06	2	88
21	0.3373	0.8378	1.78	3.70	39.43	42.98	2.98	3.06	2	90
22	0.3234	0.8267	1.78	3.66	39.17	42.90	2.98	3.06	2	92
23	0.2763	0.8150	1.58	3.62	38.56	42.81	2.81	3.05	2	94
24	0.2665	0.8092	1.42	3.59	38.44	42.76	3.07	3.05	1	95
25	0	0.8008	1.35	3.57	38.23	42.72	2.94	3.05	1	96
26	0	0.7925	1.1	3.55	38.12	42.67	3	3.05	1	97

TABLE 32

Simulated performance for census tract Queens 100

	Gender matching		Review rating		Avg profit		Avg prospect		Fulfilled requests	
Iteration	Per Iteration	Cumulative	Per Iteration	Cumulative	Per Iteration	Cumulative	Per Iteration	Cumulative	Per Iteration	Cumulative
1	1	1	5	5.00	39.65	39.65	2.93	2.93	11	11
2	1	1	4.83	4.92	39.37	39.51	3.01	2.97	11	22
3	1	1	4.76	4.87	39.24	39.43	3.1	3.01	9	31
4	1	1	4.73	4.84	39.15	39.37	2.6	2.92	9	40
5	1	1	4.68	4.81	38.87	39.29	3.15	2.96	8	48
6	1	1	4.53	4.78	38.75	39.22	2.76	2.93	7	55
7	1	1	4.51	4.75	38.64	39.15	3.12	2.95	7	62
8	0.9990	0.9990	4.2	4.69	38.4	39.08	3.04	2.96	7	69
9	0.9671	0.9959	4.19	4.65	38.18	39.00	2.89	2.95	6	75
10	0.9682	0.9936	4.17	4.62	37.55	38.91	3.18	2.97	5	80
11	0.8832	0.9869	4.11	4.59	37.41	38.82	3.02	2.97	5	85
12	0.8476	0.9788	4.07	4.56	37.31	38.74	2.75	2.96	5	90
13	0.83 28	0.9724	3.74	4.53	36.77	38.66	3.02	2.96	4	94
14	0.8177	0.9654	3.7	4.49	36.13	38.55	2.67	2.95	4	98
15	0.7963	0.9585	3.67	4.46	36.11	38.46	2.87	2.95	4	102
16	0.7524	0.9526	3.49	4.43	35.35	38.37	3.02	2.95	3	105
17	0.7212	0.9461	3.12	4.40	35.23	38.28	3.06	2.95	3	108
18	0.6616	0.9384	3.04	4.36	34.99	38.19	2.8	2.95	3	111
19	0.6109	0.9297	2.84	4.32	34.55	38.10	3.05	2.95	3	114
20	0.5956	0.9210	2.46	4.27	34.39	38.00	2.9	2.95	3	117
21	0.5573	0.9118	2.39	4.23	34.03	37.90	3.11	2.95	3	120
22	0.5489	0.9057	2.27	4.19	33.48	37.83	2.98	2.95	2	122
23	0.4871	0.8988	2.25	4.16	33.15	37.75	3.02	2.96	2	124
24	0.4833	0.8921	2.23	4.13	33.06	37.68	3.17	2.96	2	126
25	0.4125	0.8846	1.88	4.10	32.6	37.60	2.61	2.95	2	128
26	0.3886	0.8768	1.87	4.06	32.53	37.52	2.96	2.95	2	130
27	0.3841	0.8693	1.51	4.02	32.26	37.44	2.7	2.95	2	132
28	0.3723	0.8619	1.46	3.98	31.59	37.36	2.85	2.95	2	134
29	0.3406	0.8542	1.43	3.95	31.56	37.27	2.78	2.95	2	136
30	0.3159	0.8463	1.29	3.91	31.54	37.19	3.13	2.95	2	138
31	0.2333	0.8375	1.24	3.87	30.85	37.10	2.63	2.94	2	140
32	0.2219	0.8331	1.24	3.85	30.77	37.05	3.03	2.94	1	141
33	0	0.8273	1.22	3.83	30.65	37.01	2.96	2.94	1	142
34	0	0.8215	1.02	3.81	30.19	36.96	3.18	2.95	1	143

TABLE 33

Simulated performance for census tract Queens 1900

	Gender matching		Review rating		Avg profit		Avg prospect		Fulfilled requests	
Iteration	Per Iteration	Cumulative	Per Iteration	Cumulative	Per Iteration	Cumulative	Per Iteration	Cumulative	Per Iteration	Cumulative
1	1	1	4.65	4.65	41.9	41.9	2.92	2.92	9	9
2	0.9949	0.9953	4.58	4.62	40.85	41.41	2.89	2.91	8	17
3	0.9916	0.9936	4.38	4.54	40.53	41.13	2.71	2.84	8	25
4	0.9917	0.9929	4.33	4.50	40.49	41.00	3.07	2.89	6	31
5	0.9725	0.9892	4.2	4.45	39.99	40.84	2.82	2.88	6	37
6	0.9334	0.9821	3.61	4.35	39.97	40.74	2.78	2.86	5	42
7	0.8671	0.9715	3.6	4.29	39.84	40.66	2.72	2.85	4	46
8	0.8519	0.9618	3.33	4.21	39.6	40.57	2.83	2.85	4	50
9	0.8024	0.9526	3.09	4.15	38.85	40.48	2.96	2.86	3	53
10	0.7731	0.9429	2.98	4.08	38.08	40.35	3.02	2.87	3	56
11	0.7598	0.9331	2.96	4.03	36.26	40.14	2.89	2.87	3	59
12	0.7264	0.9227	2.85	3.97	35.35	39.91	2.89	2.87	3	62
13	0.6153	0.9130	2.77	3.93	34.19	39.73	2.87	2.87	2	64
14	0.5153	0.9008	2.63	3.89	34.15	39.56	2.89	2.87	2	66
15	0.4609	0.8878	2.12	3.84	32.83	39.36	2.73	2.86	2	68
16	0.3927	0.8736	2.06	3.79	32.6	39.17	2.96	2.87	2	70
17	0.3819	0.8599	1.94	3.74	32.51	38.98	3	2.87	2	72
18	0.3533	0.8461	1.73	3.68	31.54	38.78	2.85	2.87	2	74
19	0	0.8348	1.59	3.66	31.4	38.68	2.89	2.87	1	75
20	0	0.8238	1.04	3.62	31.33	38.59	3.05	2.87	1	76

TABLE 34

Simulated performance for census tract Brooklyn 1500

	Gender matching		Review rating		Avg profit		Avg prospect		Fulfilled requests	
Iteration	Per Iteration	Cumulative	Per Iteration	Cumulative	Per Iteration	Cumulative	Per Iteration	Cumulative	Per Iteration	Cumulative
1	0.9674	0.9674	4.77	4.77	44.69	44.69	2.69	2.69	9	9
2	0.9466	0.9577	4.69	4.73	44.01	44.37	2.67	2.68	8	17
3	0.8715	0.9301	4.49	4.65	42.69	43.83	2.76	2.71	8	25
4	0.8342	0.9115	3.93	4.51	41.49	43.38	2.76	2.72	6	31
5	0.8013	0.8937	3.82	4.40	41.38	43.05	2.81	2.73	6	37
6	0.7292	0.8741	3.76	4.33	40.94	42.80	2.72	2.73	5	42
7	0.7164	0.8573	3.64	4.25	40.53	42.56	2.77	2.73	5	47
8	0.6991	0.8449	3.45	4.19	40.21	42.38	2.69	2.73	4	51
9	0.6854	0.8360	3.27	4.14	39.84	42.24	3.09	2.75	3	54
10	0.6836	0.8280	3.15	4.09	39.81	42.11	2.88	2.76	3	57
11	0.5913	0.8162	3.14	4.04	39.7	41.99	2.94	2.77	3	60
12	0.5717	0.8045	3.11	3.99	39.65	41.88	3	2.78	3	63
13	0.5268	0.7960	3.08	3.97	39.43	41.80	2.71	2.78	2	65
14	0.4554	0.7858	3	3.94	38.09	41.69	2.93	2.78	2	67
15	0.4136	0.7750	2.89	3.91	37.99	41.58	2.85	2.78	2	69
16	0.3879	0.7641	2.78	3.88	37.87	41.48	3.01	2.79	2	71
17	0.3145	0.7518	2.58	3.84	37.64	41.37	2.66	2.79	2	73
18	0.2815	0.7393	2.44	3.80	36.52	41.24	3.05	2.79	2	75
19	0.265	0.7330	2.06	3.78	36.39	41.18	2.95	2.79	1	76
20	0.2222	0.7264	1.54	3.75	35.22	41.10	2.74	2.79	1	77
21	0.0754	0.7181	1.18	3.72	34.89	41.02	2.95	2.80	1	78
22	0.0707	0.7099	1.07	3.68	34.63	40.94	3.1	2.80	1	79

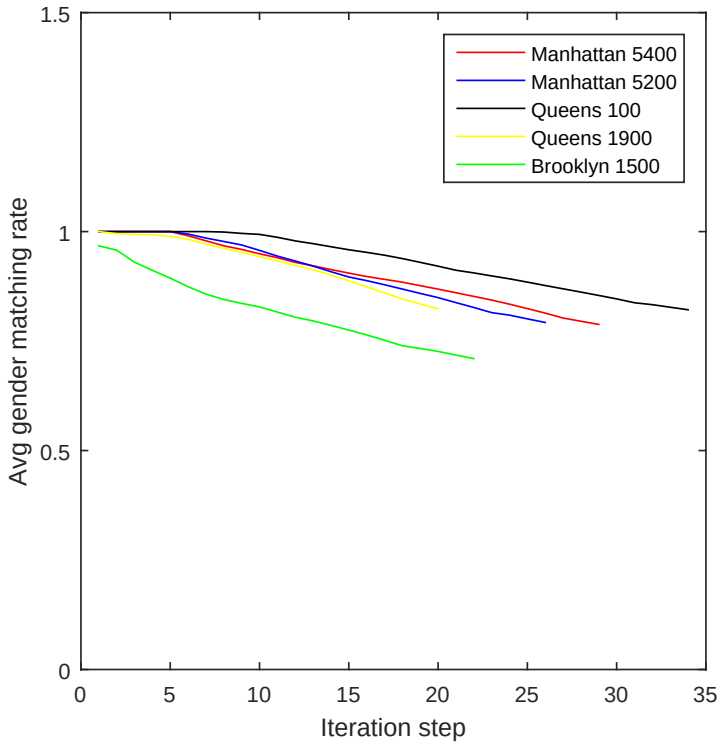

Figure 25. Cumulative gender matching rate for each iteration

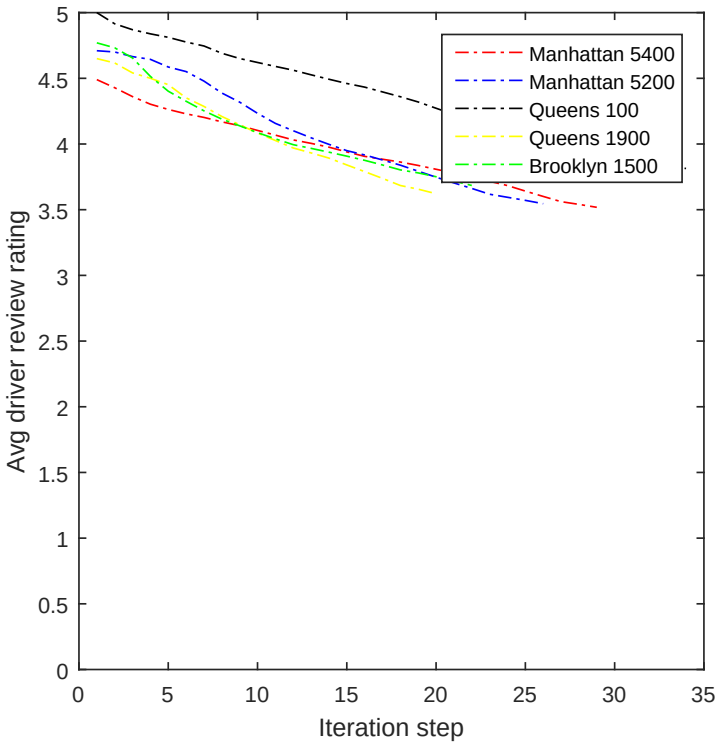

Figure 26. Cumulative review rating for each iteration

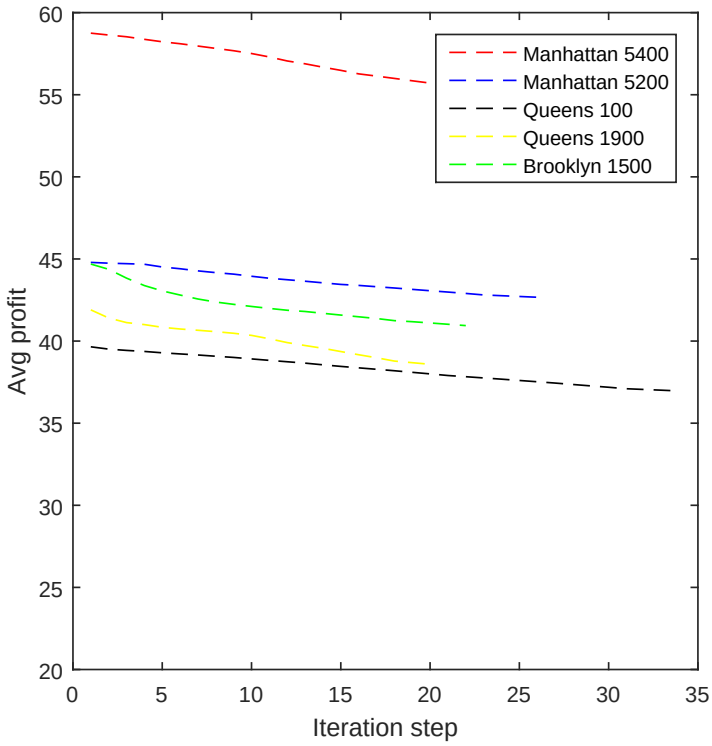

Figure 27. Cumulative average profit for each iteration

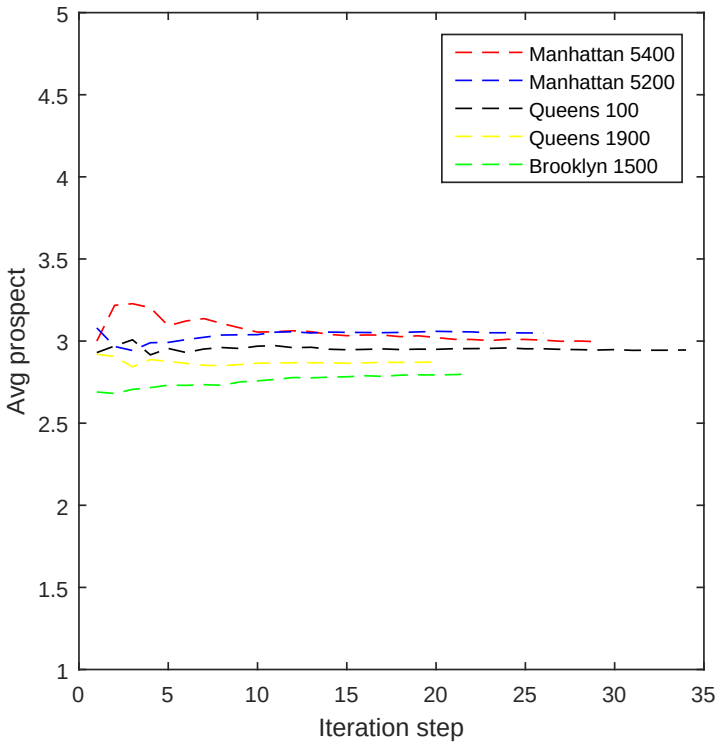

Figure 28. Cumulative average prospect for each iteration

CHAPTER 6

CONCLUSION AND FUTURE WORK

6.1 Conclusion

Collaborative ridesharing platform promises an efficient and economical transportation method via advanced mobile technologies. Most of the studies concentrate on centralized planning, aiming to maximize the system utilization. However, decision making on end-user (either driver or passenger) side has gained renewed interest recently.

In this study, we propose a two-stage approach to to solve the passenger-driver assignment problem for the centralized platform and to provide a decision-making procedure for drivers to determine bidding prices and optimal choices of passengers. In the first stage, we develop a multi-criterion many-to-many passenger-driver assignment (MCMMAP) to maximize the average gender matching rate, the average assigned drivers' review rating and the system-wide profit. A novel reformulation is introduced to solve the mixed integer non-linear programming problem (MINLP). A lexicographical solution algorithm is developed to solve the linearized equivalent problem. In the second stage, a driver's prospect maximization model is developed to help drivers decide on the optimal bidding prices for the multiplel passenger requests he/she has reveived from the first stage and ultimate decide on the optimal passenger request to accept. We show that there exists a unique global optimal solution to the problem. In addition, we develop sufficient conditions under which the driver's prospect maximization problem yields the optimal solution.

Numerically, five sets of random instances with various driver to passenger ratios in the first stage, each set of random instances have five scenarios with different combinations of utility function coefficient α and risk attitude coefficient λ. With the pre-assignment solution from the first stage, the drivers are able to identify which passenger results in the highest prospect in the second stage. Our sensitivity analysis suggests that 1) the pre-determined number of backup drivers, $\bar{D}$, will not only yield more service coverage but also create a more competitive scenario for drivers in the second stage and consequently result in driver's conservative bidding. Thus, the centralized planner needs to be careful on selecting the proper number of backup drivers; 2) as the utility coefficient α increases, the drivers tend to bid aggressively. Similar effects can be observed when the risk-related negative outcome coefficient λ decreases. However, the utility coefficient α does have larger impact due to exponential growth rate.

To further test our proposed model, we collect the New York City taxi data in April 2014. We also collected relevant demographic data to perform the statistic analysis on the taxi supply and demand forecasting model for a given urban area. Three regression models are developed. First, we explicitly examine the relationship between travel distance and total fare via a univariate linear regression model. Second, we use Poisson regression and Holt-Winter's method to forecast the demand and supply for a certain census track within given time period. Finally, with the established forecasting models, we then conduct numerical simulation with five census tracts include Manhattan 5400, Manhattan 5200, Queens 100, Queens 1900 and Brooklyn 1500. All proposed solution algorithms successfully yield quality solutions to the ridesharing assignment and auction models.

6.2 Future Research

Researchers have long emphasized the importance of mobility of physical goods with respects to transportation, material handling and inventory and storage,

to name a few. However, some have revealed (e.g., [86], [87]) that current logistics organization is not economically sustainable or environmentally friendly. Most of the current logistics networks are fragmented in terms of service type, product category and transportation purpose, each dedicated to a specific system, such as a vehicle manufacturer, a retail supply chain or an express delivery service. From the transportation perspective, modes are subjected to the specific logistics networks even if they share the same infrastructure. From the distribution and supply perspectives, the dominant paradigm focuses on vertical coordination with the well-known supply chain concept. Furthermore, the way freight are currently transported producing huge waste. The 2009 Department of Transportation freight traffic analysis [88] reports that only 60% of the capacity is utilized when trailers are traveling in the US. Globally, the transport efficacy is to be estimated as low as 10%. Besides, some of the transportation facilities and urban transportation networks are poorly designed for easing freight transportation, handling, and storage, which would create significant traffic congestion, greenhouse gas emission, and other pollution concerns [89]. The term Physical Internet was firstly introduced by *The Economist* in June 2006 [90]. Montreuil et al. [91] presented the Physical Internet as a solution to address the Global Logistics Sustainability Grand Challenge from economic, environmental and social aspects. The simulation results indicate the total roundtrip hours between Quebec and LA can be reduced by four folds under the paradigm of PI. For a comprehensive study towards Physical Internet, we refer to [92], in which the Physical Internet is defined as a consolidation of global logistics system based on physical, digital and operational interconnectivity through encapsulation, interfaces, and protocols. The functional design was studied in [92], for which the basic functional elements are listed as follows:

1. PI-container is the key elements which enable the interoperability necessary

2. PI-nodes are the necessary facilities in order to operate the Physical Internet, where π-containers' are transferred, switched, sorted and stored.

3. PI-movers are used to move the π-containers such as transporting, conveying, handling, lifting and manipulating. The main types of π-movers can be categorized as π-transporters, π-conveyors and π-handlers with respect to various functions.

With the advent of Physical Internet, it is possible to integrate Physical Internet with interconnected crowdsourcing delivery system. Similar to the Uber-like platform discussed in this dissertation, our proposed decision-making framework is able to serve as a decision-making tool for those drivers who work on such crowdsourcing delivery system.

As the sharing economy scales up in our society, it is important to examine the potential impact and implications of those crowdsourcing ridesharing platform. Specifically, its impact can be empirically examine how the entry of those vehicles into major urban areas influences traffic congestion. Leveraging on the information technology, it is possible to trace those vehicles entering different urban areas at different time, then we will be able to compare the difference in congestion in two aspects, for example, the traffic congestion before and after those vehicles that enter different urban area, or the traffic congestion between urban areas with or without ridesharing service. Obviously, ridesharing helps to increase vehicle occupancy by having more than one person in the car. Therefore the total number of cars on the road and traffic congestion decreases. Furthermore, the matching mechanism enhances the connection between passengers and drivers, thus greatly reduces the waiting time for passengers and searching/idle time for drivers.

However, ridesharing as a relatively new phenomenon, and its consequence in traffic planning still needs to be carefully investigated. One conventional approach to transportation forecasting is based on a commonly known four-step model", which was firstly introduced by Manheim [93] and expanded by Florian et al. [94]. The four-step model includes trip generation, trip distribution, mode choice and route assignment. In this dissertation, we develop the trip generation models

through Poisson regression and Holt-Winter's method. Possible improvements to these four step procedure are listed below:

1. Trip distribution is essentially a destination choice model and generates a trip matrix $T_{i,j}$ for each trip purpose utilized in the trip generation model. One commonly used approach is so called gravity model in the form: $T_{i,j} = K_i K_j T_j T_k f(C_{i,j})$. Where $T_{i,j}$ is the trips between origin i and destination j, $C_{i,j}$ is the travel cost between origin i and destination j, K_i, K_j and f are the balance factors and distance decay factor. If origin j and destination j are far from each other, it is less likely that there is interaction between them when all other conditions are equal. We can further integrate the gravity model into our decision-making support framework for the drivers, by minimizing the total distance.

2. Mode choice is the third step in travel demand forecasting process. Once the relative cost, accessibility of ridesharing versus other mode (e.g., public transit, drive) is examined, a logit regression model can be use to predict the probability of choosing one transportation mode over others.

3. Route assignment is the last step in trip forecasting model. Once the previous three models are generated, the driver equilibrium model can be applied for advanced transportation planning analysis.

REFERENCES

[1] N. D. Chan and S. A. Shaheen, *Ridesharing in north america: Past, present, and future,* Transport Reviews 32, 93–112 (2012).

[2] U. S. C. Bureau, *American community survey 1-year estimates,* (2011).

[3] A. Amey, J. Attanucci, and R. Mishalani, *Real-time ridesharing: opportunities and challenges in using mobile phone technology to improve rideshare services,* Transportation Research Record: Journal of the Transportation Research Board pp. 103–110 (2011).

[4] N. Agatz, A. Erera, M. Savelsbergh, and X. Wang, *Optimization for dynamic ride-sharing: A review,* European Journal of Operational Research 223, 295–303 (2012).

[5] S. N. Parragh, K. F. Doerner, and R. F. Hartl, *A survey on pickup and delivery problems,* Part II: Transportation between pickup and delivery locations, to appear: Journal für Betriebswirtschaft (2007).

[6] F. Alt, A. S. Shirazi, A. Schmidt, U. Kramer, and Z. Nawaz, *Location-based crowdsourcing: extending crowdsourcing to the real world,* in *Proceedings of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries,* (ACM, 2010), pp. 13–22.

[7] K. Ghoseiri, A. E. Haghani, M. Hamedi, and M. Center, *Real-time rideshare matching problem* (Mid-Atlantic Universities Transportation Center Berkeley, 2011).

[8] A. Amey, *Integrating information on ridesharing opportunities,* TDM Review 17 (2010).

[9] S. Heinrich, *Implementing real-time ridesharing in the san francisco bay area,* Master, San Jose State University (2010).

[10] N. Agatz, A. L. Erera, M. W. Savelsbergh, and X. Wang, *Dynamic ride-sharing: A simulation study in metro atlanta,* Procedia-Social and Behavioral Sciences 17, 532–550 (2011).

[11] C.-C. Tao and C.-Y. Chen, *Heuristic algorithms for the dynamic taxipooling problem based on intelligent transportation system technologies,* in *Fuzzy Systems and Knowledge Discovery, 2007. FSKD 2007. Fourth International Conference on,* , vol. 3 (IEEE, 2007), vol. 3, pp. 590–595.

[12] D. Kahneman and A. Tversky, *Prospect theory: An analysis of decision under risk,* in *Handbook of the fundamentals of financial decision making: Part I,* (World Scientific, 2013), pp. 99–127.

[13] M. Furuhata, M. Dessouky, F. Ordóñez, M.-E. Brunet, X. Wang, and S. Koenig, *Ridesharing: The state-of-the-art and future directions,* Transportation Research Part B: Methodological 57, 28–46 (2013).

[14] R. W. Hall and A. Qureshi, *Dynamic ride-sharing: Theory and practice,* Journal of Transportation Engineering 123, 308–315 (1997).

[15] M. Ben-Akiva and T. J. Atherton, *Methodology for short-range travel demand predictions: Analysis of carpooling incentives,* Journal of Transport Economics and Policy pp. 224–261 (1977).

[16] K. Train, *A validation test of a disaggregate mode choice model,* Transportation Research 12, 167–174 (1978).

[17] I. P. Levin, *Measuring tradeoffs in carpool driving arrangement preferences,* Transportation 11, 71–85 (1982).

[18] E. Ferguson, K. Hodge, and K. Berkovsky, *Psychological benefits from vanpooling and group composition,* Transportation 21, 47–69 (1994).

[19] H.-J. Huang, H. Yang, and M. G. Bell, *The models and economics of carpools,* The annals of regional science 34, 55–68 (2000).

[20] K. Washbrook, W. Haider, and M. Jaccard, *Estimating commuter mode choice: A discrete choice analysis of the impact of road pricing and parking charges,* Transportation 33, 621–639 (2006).

[21] A. Amey, *A proposed methodology for estimating rideshare viability within an organization, applied to the mit community,* in *TRB Annual Meeting Procediings,* (2011), pp. 1–16.

[22] R. Baldacci, V. Maniezzo, and A. Mingozzi, *An exact method for the car pooling problem based on lagrangean column generation,* Operations Research 52, 422–439 (2004).

[23] R. W. Calvo, F. de Luigi, P. Haastrup, and V. Maniezzo, *A distributed geographic information system for the daily car pooling problem,* Computers & Operations Research 31, 2263–2278 (2004).

[24] N. H. M. Wilson and N. J. Colvin, *Computer control of the Rochester dial-a-ride system*, 77 (Massachusetts Institute of Technology, Center for Transportation Studies, 1977).

[25] N. H. Wilson, R. W. Weissberg, and J. Hauser, *Advanced dial-a-ride algorithms research project,* Tech. rep. (1976).

[26] N. H. Wilson, J. Sussman, H.-K. Wong, and T. Higonnet, *Scheduling algorithms for a dial-a-ride system* (Massachusetts Institute of Technology. Urban Systems Laboratory, 1971).

[27] H. N. Psaraftis, *A dynamic programming solution to the single vehicle many-to-many immediate request dial-a-ride problem,* Transportation Science 14, 130–154 (1980).

[28] H. N. Psaraftis, *An exact algorithm for the single vehicle many-to-many dial-a-ride problem with time windows,* Transportation science 17, 351–357 (1983).

[29] R. B. Dial, *Autonomous dial-a-ride transit introductory overview,* Transportation Research Part C: Emerging Technologies 3, 261–275 (1995).

[30] J. Yang, P. Jaillet, and H. Mahmassani, *Real-time multivehicle truckload pickup and delivery problems,* Transportation Science 38, 135–148 (2004).

[31] S. Nittel, M. Duckham, and L. Kulik, *Information dissemination in mobile ad-hoc geosensor networks,* in *International Conference on Geographic Information Science,* (Springer, 2004), pp. 206–222.

[32] S. Winter and S. Nittel, *Ad hoc shared-ride trip planning by mobile geosensor networks,* International Journal of Geographical Information Science 20, 899–916 (2006).

[33] X. Xing, T. Warden, T. Nicolai, and O. Herzog, *Smize: a spontaneous ride-sharing system for individual urban transit,* in *German Conference on Multiagent System Technologies,* (Springer, 2009), pp. 165–176.

[34] A. Kleiner, B. Nebel, and V. A. Ziparo, *A mechanism for dynamic ride sharing based on parallel auctions,* in *IJCAI,* , vol. 11 (2011), vol. 11, pp. 266–272.

[35] J. Howe, *The rise of crowdsourcing,* Wired magazine 14, 1–4 (2006).

[36] G. D. Saxton, O. Oh, and R. Kishore, *Rules of crowdsourcing: Models, issues, and systems of control,* Information Systems Management 30, 2–20 (2013).

[37] D. Brabahan, *Crowdsourcing as a model for problem solving,* The International Journal of Research into New Media Technologies 14, 75–90 (2008).

[38] V. Carbone, A. Rouquet, and C. Roussat, *Carried away by the crowd: what types of logistics characterise collaborative consumption,* in *1st International Workshop on Sharing Econom, Utrecht, Netherlands,* (2015).

[39] S. Lee, Y. Kang, and V. V. Prabhu, *Smart logistics: distributed control of green crowdsourced parcel services,* International Journal of Production Research 54, 6956–6968 (2016).

[40] A. Doan, R. Ramakrishnan, and A. Y. Halevy, *Crowdsourcing systems on the world-wide web,* Communications of the ACM 54, 86–96 (2011).

[41] K. Suh, T. Smith, and M. Linhoff, *Leveraging socially networked mobile ict platforms for the last-mile delivery problem,* Environmental science & technology 46, 9481–9490 (2012).

[42] A. Barr and J. Wohl, *Exclusive: Walmart may get customers to deliver packages to online buyers,* REUTERS–Business Week (2013).

[43] A. M. Arslan, N. Agatz, L. Kroon, and R. Zuidwijk, *Crowdsourced deliverya dynamic pickup and delivery problem with ad hoc drivers,* Transportation Science (2018).

[44] B. Ferris, K. Watkins, and A. Borning, *Onebusaway: results from providing real-time arrival information for public transit,* in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems,* (ACM, 2010), pp. 1807–1816.

[45] F. Filippi, G. Fusco, and U. Nanni, *User empowerment and advanced public transport solutions,* Procedia-Social and Behavioral Sciences 87, 3–17 (2013).

[46] H. W. Kuhn, *The hungarian method for the assignment problem,* Naval research logistics quarterly 2, 83–97 (1955).

[47] J. Munkres, *Algorithms for the assignment and transportation problems,* Journal of the society for industrial and applied mathematics 5, 32–38 (1957).

[48] H. Zhu, D. Liu, S. Zhang, Y. Zhu, L. Teng, and S. Teng, *Solving the many to many assignment problem by improving the kuhn–munkres algorithm with backtracking,* Theoretical Computer Science 618, 30–41 (2016).

[49] I. Litvinchev, M. Mata, S. Rangel, and J. Saucedo, *Lagrangian heuristic for a class of the generalized assignment problems,* Computers & Mathematics with Applications 60, 1115–1123 (2010).

[50] I. Litvinchev, S. Rangel, and J. Saucedo, *A lagrangian bound for many-to-many assignment problems,* Journal of Combinatorial Optimization 19, 241–257 (2010).

[51] E. H. Durfee, J. C. Boerkoel, and J. Sleight, *Using hybrid scheduling for the semi-autonomous formation of expert teams,* Future Generation Computer Systems 31, 200–212 (2014).

[52] C.-L. Hwang, S. R. Paidy, K. Yoon, and A. S. M. Masud, *Mathematical programming with multiple objectives: A tutorial,* Computers & Operations Research 7, 5–31 (1980).

[53] J. L. Ringuest, *Multiobjective optimization: behavioral and computational considerations* (Springer Science & Business Media, 2012).

[54] R. E. Steuer, *Multiple criteria optimization: theory, computation, and applications* (Wiley, 1986).

[55] R. T. Marler and J. S. Arora, *Survey of multi-objective optimization methods for engineering,* Structural and multidisciplinary optimization 26, 369–395 (2004).

[56] J. Andersson, *A survey of multiobjective optimization in engineering design,* Department of Mechanical Engineering, Linktjping University. Sweden (2000).

[57] Y. Y. Haimes, *On a bicriterion formulation of the problems of integrated system identification and system optimization,* IEEE transactions on systems, man, and cybernetics 1, 296–297 (1971).

[58] J. Von Neumann and O. Morgenstern, *Theory of games and economic behavior (commemorative edition)* (Princeton university press, 2007).

[59] J. D. Hey and C. Orme, *Investigating generalizations of expected utility theory using experimental data,* Econometrica: Journal of the Econometric Society pp. 1291–1326 (1994).

[60] N. Barberis and M. Huang, *Stocks as lotteries: The implications of probability weighting for security prices,* American Economic Review 98, 2066–2100 (2008).

[61] J. Sydnor, *(over) insuring modest risks,* American Economic Journal: Applied Economics 2, 177–99 (2010).

[62] C. Camerer, L. Babcock, G. Loewenstein, and R. Thaler, *Labor supply of new york city cabdrivers: One day at a time,* The Quarterly Journal of Economics 112, 407–441 (1997).

[63] S. Harding, M. Kandlikar, and S. Gulati, *Taxi apps, regulation, and the market for taxi journeys,* Transportation Research Part A: Policy and Practice 88, 15–25 (2016).

[64] J. Wirtz and C. Tang, *Uber: Competing as market leader in the us versus being a distant second in china,* in *SERVICES MARKETING: People Technology Strategy,* (2016), pp. 626–632.

[65] N. J. Garber and L. A. Hoel, *Traffic and highway engineering* (Cengage Learning, 2014).

[66] R. Turvey, *Some economic features of the london cab trade,* The Economic Journal 71, 79–92 (1961).

[67] R. B. Coffman and C. Shreiber, *The economic reasons for price and entry regulation of taxicabs (comment and rejoinder),* Journal of Transport Economics and Policy pp. 288–304 (1977).

[68] P. S. Dempsey, *Taxi industry regulation, deregulation & (and) reregulation: The paradox of market failure,* Transp. LJ 24, 73 (1996).

[69] C. Gaunt, *The impact of taxi deregulation on small urban areas: some new zealand evidence,* Transport Policy 2, 257–262 (1995).

[70] A. Marell and K. Westin, *The effects of taxicab deregulation in rural areas of sweden,* Journal of Transport Geography 10, 135–144 (2002).

[71] A. Kumar and D. M. Levinson, *Specifying, estimating and validating a new trip generation model: Case study in montgomery county, maryland,* (1993).

[72] W. A. ONeill and E. Brown, *Long-distance trip generation modeling using ats,* in *conference Personal Travel: The Long and Short of It, Transportation Research Board, June,* (1999).

[73] B. Schaller, *A regression model of the number of taxicabs in us cities,* Journal of Public Transportation 8, 4 (2005).

[74] A. Mousavi, J. M. Bunker, and B. Lee, *A new approach for trip generation estimation for use in traffic impact assessments,* in *25th ARRB Conference Proceedings,* (ARRB Group Ltd, 2012).

[75] G. Corpuz, *Public transport or private vehicle: factors that impact on mode choice,* in *30th Australasian Transport Research Forum,* (2007), p. 11.

[76] Y.-C. Chang, *Factors affecting airport access mode choice for elderly air passengers,* Transportation research part E: logistics and transportation review 57, 105–112 (2013).

[77] J. Ben-Edigbe and R. Rahman, *Multivariate school travel demand regression based on trip attraction,* World Acad. Sci. Engg. Technol 66, 1695–1699 (2010).

[78] R. Ewing, W. Schroeer, and W. Greene, *School location and student travel analysis of factors affecting mode choice,* Transportation Research Record: Journal of the Transportation Research Board pp. 55–63 (2004).

[79] E. Rosenthal, *Gams-a users guide,* in *GAMS Development Corporation,* (Citeseer, 2008).

[80] R. E. Rosenthal, *Gams–a user's guide,* (2004).

[81] P. Koster, E. Kroes, and E. Verhoef, *Travel time variability and airport accessibility,* Transportation Research Part B: Methodological 45, 1545–1559 (2011).

[82] J. Ord, *Handbook of the poisson distribution,* Journal of the Operational Research Society 18, 478–479 (1967).

[83] P. S. Kalekar, *Time series forecasting using holt-winters exponential smoothing,* Kanwal Rekhi School of Information Technology 4329008, 1–13 (2004).

[84] G. P. Nason, *Stationary and non-stationary time series,* Statistics in Volcanology. Special Publications of IAVCEI 1, 000–000 (2006).

[85] P. R. Winters, *Forecasting sales by exponentially weighted moving averages,* Management science 6, 324–342 (1960).

[86] R. Dekker, J. Bloemhof, and I. Mallidis, *Operations research for green logistics–an overview of aspects, issues, contributions and challenges,* European Journal of Operational Research 219, 671–679 (2012).

[87] P. R. Murphy and R. F. Poist, *Green perspectives and practices: a comparative logistics study,* Supply chain management: an international journal 8, 122–131 (2003).

[88] U. C. Bureau, *Statistical Abstract of the United States 2009 (Hardcover)* (Government Printing Office, 2008).

[89] T. L. Magnanti and R. T. Wong, *Network design and transportation planning: Models and algorithms,* Transportation science 18, 1–55 (1984).

[90] P. Markillie, *The physical internet: A survey of logistics* (Economist Newspaper, 2006).

[91] B. Montreuil, R. D. Meller, and E. Ballot, *Towards a physical internet: the impact on logistics facilities and material handling systems design and innovation,* (2010).

[92] B. Montreuil, *Toward a physical internet: meeting the global logistics sustainability grand challenge,* Logistics Research 3, 71–87 (2011).

[93] M. L. Manheim, *Fundamentals of transportation systems analysis*, vol. 1 (Mit Press Cambridge, MA, 1979).

[94] M. Florian, M. Gaudry, and C. Lardinois, *A two-dimensional framework for the understanding of transportation planning models,* Transportation Research Part B: Methodological 22, 411–419 (1988).

CURRICULUM VITAE

NAME: Peiyu Luo

ADDRESS: Department of Industrial Engineering
University of Louisville
Louisville, KY 40292

EDUCATION: B.E. Environment Engineering
Jiangnan University (China)
2008
M.S. Industrial Engineering
University of Louisville
2013

PREVIOUS RESEARCH: Logistics
Operations Research
Optimization

AWARDS: University Fellowship 2011 - 2012
Alpha Pi Mu (Industrial Engineering Honor Society)